White Paper Report

Report ID: 104546

Application Number: HD5143311

Project Director: Alan Liu (ayliu@english.ucsb.edu)

Institution: University of California, Santa Barbara

Reporting Period: 9/1/2011-9/30/2012

Report Due: 12/31/2012

Date Submitted: 12/21/2012

FRIENDING THE HUMANITIES KNOWLEDGE BASE:
EXPLORING BIBLIOGRAPHY AS SOCIAL NETWORK IN RoSE

Alan Liu (ayliu@english.ucsb.edu), Rama Hoetzlein, Rita Raley, Ivana Anjelkovic, Salman Bakht, Joshua Dickinson, Michael Hetrick, Andrew Kalaidjian, Eric Nebeker, Dana Solomon, and Lindsay Thomas

University of California, Santa Barbara

Figure 1: RoSE home page. (See Appendix A for larger screenshots from RoSE.)

WHAT IF?

• What if bibliographies of past authors and works could be modeled as a dynamic, evolving society linked to today's scholars and students?

• What if scholars and students could add data about biographical, historical, and intellectual relationships to the bibliographical entries, thus using present-day "crowdsourcing" to make more socially meaningful the crowds of history?

• What if humanities resources--sometimes with non-conformant metadata from the distant past--could especially benefit from this process?

• What if we could change the nature of initial research from "searching" to participatory "making"?

• What if visualizations could help us actively "storyboard" intellectual movements and not just spectate them?

• And, recursively, what if such a system for active learning with interactive technology could mirror the way the system developers themselves collaborated--integrating the humanities, arts, and engineering to explore the humanistic issues in technological problems, and the technological issues in humanities problems?

1. INTRODUCTION AND SUMMARY

We report in this paper on a project we advanced from an initial prototype to beta stage in 2011-12 with a NEH Digital Humanities Start-up Grant (Level 2). We aim not just to narrate grant objectives, activities, and results but also to surface some of the larger digital humanities issues--inextricably humanistic and technological, theoretical and practical--that we engaged.

The project is called RoSE (Research-oriented Social Environment), an online knowledge exploration environment for humanities scholars and students developed in the Ruby on Rails programming environment on top of a MySQL database. Accessed through a Web site (http://rose.english.ucsb.edu), the system includes the following main content and interface features: • an extensive set of bibliographical metadata (but no full texts) machine-harvested from Project Gutenberg, YAGO, and SNAC (Social Networks & Archival Contexts); • an initial set of user-entered metadata (including "relationships" and "keywords") added to the pre-existing data; • a user interface with search and editing functionality modeled as a social network site with "profile pages" for each author, work, and user; • interactive visualizations in several styles to facilitate navigation and understanding; • "history"-tracking and "collections"; • "storyboards" to shape visual arguments; • and user documentation, including a "Quick Start Guide" and demo video.

ACCESSING RoSE

Explore RoSE (rose.english.ucsb.edu) as a guest user by leaving the login fields blank and clicking "login." Or request access as a registered user with a profile page who is able to add to our knowledge base (contact: ayliu@english.ucsb.edu) Currently, the RoSE beta is open on a limited basis (by request) for registration from scholars, teachers, students, and others.

The basic idea of RoSE is that a bibliography of humanities-related knowledge can be modeled as an evolving "social network" of people and works. Bibliography thus acquires a social face, becoming not a set of "records" but a participatory network of relationships between, for example, an author's colleagues, friends, lovers, imitators, critics, and later scholars and students (or a work's similar influences on and relationships with other works). We position RoSE as a contextual discovery tool for the formative stages of learning about a topic. When beginning to research an author, work, or idea, users (our target audiences are undergraduates, graduate students, and more advanced scholars) can explore RoSE to find clusters and pathways of relationships situating their topic in its intellectual context; and they can add entries and relationships as part of the very process of advancing their understanding of that context.

The larger issues addressed by RoSE are alluded to in our epigraph's "what if's?" RoSE's solutions (or, in cases of partial implementation, approaches) to these issues are our contribution to some of the major concerns shared across the digital humanities today, including: • how to combine the newer social network paradigm of collective knowledge with older paradigms of collected knowledge (e.g., resources in a library); • how to use bottom-up crowdsourcing in concert with top-down expertise; • how to make "messy" humanities data (especially of historical vintage) comport itself like the crisp, well-behaved metadata needed for computational systems; • how to design a participatory system that encourages users to learn by "making" and not just "finding" (thus shifting the "do you have to be a builder?" issue in digital humanities to its proper locus: users); • how to visualize network relations in ways that transcend passive spectacle to achieve active understanding; • and how to channel the collaboration of digital humanities developers themselves--overcoming the coding vs. interpreting, or "hack vs. yack," divide (Cecire; Koh)--into their products as a robust bias toward technical-and-interpretive solutions. (See "Reflections" at the end of this paper.)

RoSE was first prototyped by the University of California system's Transliteracies Multicampus Research Group (2005-2010; PI Alan Liu). For the NEH Digital Humanities Start-up Grant in 2011-12, we moved RoSE forward to a beta stage by expanding the project's knowledge base with additional machine-harvested bibliographic metadata (especially from the SNAC project); evolving our visualizations to facilitate navigation and interactivity; and running a limited set of "use-scenario" tests (including in an undergraduate course). These goals were accompanied by the ongoing task of improving our user interface and general usability. Midway through the grant year, we also added the significant new goal of a "storyboard" system giving users the ability to sketch knowledge gathered from RoSE as a shareable argument or narrative.

We met these goals through individual and team work paced by weekly or biweekly meetings (supplemented by separate work meetings specifically of programmers). A bonus of these meetings--which continued the spirit of the Transliteracies "Blue Sky" developer meetings in the predecessor stage of the RoSE project--was that they included graduate-level discussions of technical issues with surprising humanistic depths and, reciprocally, humanistic issues with unsuspected technical interest.

At the end of the NEH grant period, we concluded with a fully functional beta. Currently, anyone can passively access RoSE through a guest account; and we are beginning selectively to invite and approve requests for active user status (registered users with editing permissions) from the national and international academic community. Users will help us further explore our "what if" problems and give us feedback to improve our beta further in anticipation of a possible next implementation stage of the project.

Project participants included members of the UC Santa Barbara English Department and Media Arts & Technology (MAT) program:

DIRECTOR:
Alan Liu (Professor, English Dept.)

CO-DIRECTORS:
Rama Hoetzlein (Ph.D., Media Arts & Technology Program; Project Scientist, English Dept.)
Rita Raley (Associate Professor, English Dept.)

RESEARCH ASSISTANTS
(graduate students or recent Ph.D.'s):
Ivana Andjelkovic (MAT)
Salman Bakht (MAT)
Joshua Dickinson (MAT)
Michael Hetrick (MAT)
Paul Jacobs (MAT)
Andrew Kalaidjian (English)
Eric Nebeker (English)
Dana Solomon (English)
Lindsay Thomas (English)
[Paul Jacobs (MAT), research assistant in fall 2011]

We gratefully acknowledge prior work on the RoSE prototype in the Transliteracies initiative by a large team of graduate students (a few of whom continued into the NEH project) from UC Santa Barbara's Media Arts & Technology program and the English, History, Information Studies, Literature, Visual Studies departments of the following University of California campuses: UCLA, UC Santa Barbara, UC Santa Cruz, and UC Irvine.[1]

---

[1] For a listing of the RoSE team during the project's inception in the Transliteracies initiative, see http://transliteracies.english.ucsb.edu/post/research-project/rose/developmentteam.

2. PROJECT BACKGROUND

A. The RoSE Concept

Underlying RoSE is the question: what do people want when they begin online research on a topic? Our hypothesis is that people do not necessarily want to go first to a book, document, or other knowledge artifact (a bibliographical approach) or to another person (a social-network approach)--though they will take either. More ideal is an environment that allows people to seek out persons and works in the context of interrelationships of authorship, influence, sponsorship, commentary, criticism, censorship, remediation, etc., as well as of biographical and historical affiliations. In such a knowledge environment, there would be no works sitting in virtual libraries as opposed to people joining communities. Instead, authors, editors, readers, researchers, students, and the works they engage would be interlinked in a single "social-document-graph" fusing the paradigm of a document archive with that of the "social graph" familiar to social network sites and Web 2.0. People and works would be "friends," and many other things too, with other people and works.

On this main hypothesis, we built RoSE as a Web-accessible humanities knowledge base of bibliographical entries (names, classificatory keywords, and other metadata for authors and works, but not full-text works themselves) that can be explored as an authorial, reception, biographical, and scholarly "social network." The primary goal of the knowledge base is to serve as what we call a "contextual discovery system" assisting in the crucial first stages of research and learning--i.e., the formative stage when scholars or students familiarizing themselves with a new topic need to locate resources not in isolation but in relationships suggesting the extent, density, adjacencies, and rough shape of an idea or movement. "Where do I start?" "Where does that lead?" "What else should I know?" and "How might all that give me an idea?" are the sorts of questions RoSE is designed to answer.

But we had a secondary hypothesis too: that a contextual discovery system can naturally also be an understanding system--one that leads beyond locating resources in context to developing theses about the logic of that context--if "discovering" can also be "making" or "constructing." (An analogy from a generation ago might have been learning about radios by building a crystal radio receiver. Today, an updated analogy might be learning programming by making a little game in the MIT Media Lab's Scratch visual programming environment.) We thus wanted RoSE not only to come pre-stocked with metadata for authors and works (machine-harvested from existing textbases and knowledge bases) but also to allow users to "thickly describe" that metadata (borrowing a concept from cultural anthropologist Clifford Geertz) to make it more socially understandable.

Like readers of print media who highlight, underscore, and annotate in the margins, users should be able to engage in thick description as part of the reinforcing loop that is the process of learning about a topic. Taking advantage of new media, their interventions should then add to the common knowledge base for all to share.

On this secondary hypothesis, we built RoSE to be a user-editable and expandable knowledge base. Users can add entries and keywords, and make new relationships between entries. Bibliographical or prosopographical descriptions thus bloom with a social matrix of supplementary descriptions of the sort: "collaborator of," "enemy of," "lover of," "reader of," "censor of," etc.

Finally, we had a third hypothesis: the initial research acts of discovering and understanding are really only complete when they lead to the shaping of proto-arguments that can be shared (i.e., ideas clarified to be "told" to oneself or to someone else). By "proto-arguments," we mean the kind of proto-narratives characteristic of intellectual history--e.g., arguments about the "rise," "decline," "clash," "convergence," "change," and other actions of ideas that may not have much initial depth in themselves but are the necessary scaffolding on which to create more nuanced interpretations.

On this third hypothesis, we built RoSE to allow users to save "histories" and "collections" of items they find in the system; and then to use these to populate a visual "storyboard" canvas where they can filter, shape, draw, and annotate to sketch out the rough shape of an argument or narrative. These storyboards are exportable and shareable as XML files.

In this fashion, RoSE migrates the digital humanities paradigms of the digital "library," "archive," and "collection" to a social-network model that is both more native to the Web 2.0 view of the interconnectedness of knowledge agents and objects and closer to recent scholarly understandings of the "sociology of texts," the "new sociology of literature," and collective user or reader activity (McKenzie; English; Bérubé et al.). Inversely, RoSE moves the social-network paradigm closer to an older bibliographical model accommodating historical knowledge, thus enriching presentist Web 2.0 practices with the sense of history that is one cardinal characteristic of the humanities.

We imagine several specific uses for the context discovery, understanding, and argument-shaping capabilities of RoSE, ranging from pedagogy to advanced research. But one use--student research--can illustrate. Imagine an instructor introducing a humanistic "movement"--i.e., the kind of loose aggregate of people and works that scholars traditionally describe in such overlapping terms as group, circle, period, trend, or theme. Examples include Colonialism, Industrialism, Romanticism, Marxism, Dadaism, the Harlem

Renaissance, or Cyberpunk. The instructor could use RoSE to launch students into explorative initial research. She would offer initial suggestions of persons or works to follow. The students--social actors, and not just observers, in the system--would then immersively construct their understanding of the movement. They would navigate and visualize RoSE's existing knowledge base of relations between people and works. (Links in RoSE also help access additional online information or full texts elsewhere on the Web.) The heart of the assignment would then be for students to add their own thickly described connections between persons and/or works as they learn. For example, if they learn that Marcel Duchamp was a "creator of" a found-art object (like a urinal) and an "associate of" Surrealism, they might add additional relation types and connections that help define what "creator" or "associate" actually means in this case (questions that for Duchamp demand thick description). The payoff is a learning experience in which persons and documents come to social "life" in relation to each other and to the students themselves through profile pages and network visualizations witnessing the impact of the students' research on the system. Adding a relation between William Wordsworth and William Gibson, for instance, would result in the two appearing on each other's profile page, as well as on the profile page of the student researcher. The movements of "Romanticism" and "Cyberpunk" would thus come alive in a new connection mediated by the social presence of the student.
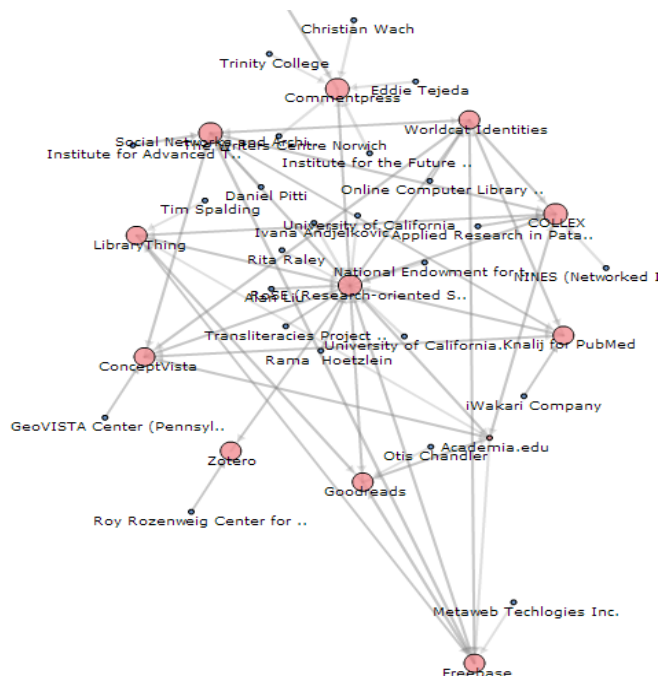
B. Environmental Scan

Figure 2: Environmental scan for RoSE (visualized in RoSE)

RoSE's innovation may best be explained by comparison with other projects studied by Transliteracies when it first prototyped RoSE. **Appendix B** lists a selection of comparison projects on which Transliteracies produced reports or papers. The most relevant include:

(i) *Advanced Humanities Online Reading, Bibliographical, and Research Environments*. Collex, Open Journal Systems (with "reading tools" extensions), PreE, Zotero, etc. focus on finding, accessing, manipulating, editing, or annotating documents--sometimes also allowing for on-demand use of data-mining tools such as those from TAPoR and SEASR. Importantly, some of these environments now borrow Web 2.0 social-network practices--e.g., folksonomical tagging and shared "bookshelves." Yet few integrate such features with their main functions, which remain bibliographical. Notable exceptions include CommentPress (used by MediaCommons), which encourages reading communities to form around parts of documents.

(ii) *Social Book Cataloging & Social Academic Networks*. Examples of social book cataloging sites include Goodreads, Visual Bookshelf (in Facebook), Living Social (Books), LibraryThing, and weRead. An example of a social academic network is Academia.edu. Where advanced humanities document environments like Collex or PreE emphasize working with documents, and add social experience on the side, social book cataloging and academic networks emphasize working with social or group identity, and add document functionality on the side. In most social book cataloging sites, for instance, books are inert objects that function like badges of the user's tastes and interests. In Academia.edu, documents can be linked and uploaded, but are also inert appendages to academic identities.

(iii) *Closest Analogues to RoSE Discovered Prior to Project Start*. The prototype for the Quanta project, developed by our co-director Rama Hoetzlein, influenced our interest in exploring graphs as an organizing paradigm, although Rose diverges from it in its objects of study, underlying representation, and method of presentation. Three of the closest analogues we discovered during the Transliteracies prototype stage for RoSE prior to the start of the NEH grant were ConceptVista, WorldCat Identities, and SNAC (Social Networks and Archival Context). ConceptVista is a research environment for geoscience that allows users to visualize and explore relations between researchers, documents, datasets, and tools. Though not entirely commensurable with humanities scholarship (which does not cite datasets and tools, but which does cite historical materials), ConceptVista is intriguing for RoSE because its operational unit of knowledge is a person linked to documents, tools, and datasets (or, put inversely, a document, tool, or dataset linked to persons). WorldCat Identities is an interface for the WorldCat network of library content and services that allows users to see information about a particular "identity" (a knowledge producer) in a one-Web-page layout. Of special interest to RoSE is the way a WorldCat Identities page features automatically-generated data--including a

publication timeline, bibliography, "related" identities, and "associated subjects." The whole amounts to something like a social-network "profile page." The difference from RoSE is that the underlying paradigm of WorldCat Identities constrains users to the role of receivers of information based on existing library metadata (entered by experts using controlled vocabularies). SNAC is a large, impressive initiative begun after RoSE was in prototype. Taking as its starting point the "finding aids" of library and archival collections, SNAC uses the EAC-CPF metadata standard (Encoded Archival Context--Corporate Bodies, Persons, and Families) to "unlock" the descriptions of persons from bibliographical data for use in new ways. A prototype user interface released in December 2010 shows the potential to allow users to see the social relation between an author (or other knowledge-producer) and related identities, documents, and collections. SNAC subsequently developed visualizations. While SNAC overlaps in concept with RoSE, it too is based on an underlying library paradigm in which the user has passive access to contexts generated from expert-created metadata.

Like ConceptVista (and unlike WorldCat Identities and SNAC), RoSE integrates the social-network and library paradigms. Like Worldcat Identities and SNAC (and unlike ConceptVista), RoSE applies the social-network model to historical materials. A scholar can link his or her profile page, for instance, to a dead author, even as that dead author's profile page changes in the relations it shows to others. The whole society of knowledge (including recent research) evolves.

Since the start of the RoSE project, new comparison projects (or ones newly known to us) came to attention. These include: • digital humanities projects that map or visualize past authors as a social network (e.g., Mapping the Republic of Letters; The Crowded Page; the JIT hypergraph network visualizations in the Women Writers Project); • bibliographical network visualization systems in the sciences (e.g., Knalij for PubMed). Additionally, we corresponded or presented at conferences organized by other digital humanities projects currently developing social-network style visualizations for their resources--e.g., The Orlando Project and The Yaddo Circles project.

C. History of Project

RoSE first arose in prototype form as the final experiment during 2008-10 of the University of California Multi-campus Research Group on online reading called "Transliteracies" (funded for 2005-2010 by the UC Office of the President; PI, A. Liu; Web site: http://transliteracies.english.ucsb.edu). The overall Transliteracies initiative involved a total of 61 faculty, 40 graduate students, and 7 industry professionals from 31 programs and 35 institutions contributing to conferences, workshops, and a "Social Computing in 2020" innovation contest. In its first years, the project's working groups (on history of reading, new reading interfaces, and social computing) conducted an extensive environmental scan that produced 78 online research reports and 7 fuller research papers (see the project site). In its final two years, the project then focused on an under-examined area: social-computing technologies that could bridge integrally between humanities and Web 2.0 online environments. Working with graduate students in the humanities, media arts, and engineering in biweekly developer meetings (which doubled as de facto seminars on fascinating blends of technical and conceptual issues), Transliteracies devoted a year and a half in 2008-10 to creating the RoSE prototype, which was presented for critique at a February 2010 "design charrette" attended by sixteen UC faculty and respondents from industry and other universities.

After the conclusion of Transliteracies, Alan Liu (with Rita Raley) applied for and received a NEH Digital Humanities Start-up Grant (Level II, $50,000) to continue work on RoSE. Liu served as director, while Raley and Rama Hoetzlein, who had previously helped develop RoSE in its Transliteracies phase while completing his Ph.D. in Media Arts & Technology, served as co-directors. Hoetzlein was also Project Scientist for the project and led the programming team. Research assistants for the project were recruited from the UC Santa Barbara English Department and Media Arts & Technology Program.

3. PROJECT OBJECTIVES, ACTIVITIES, AND RESULTS IN NEH GRANT PERIOD

For our NEH grant period (Sept. 1, 2011, to Sept. 30, 2012), we originally proposed three development objectives to address weaknesses in the RoSE prototype and also to evaluate the system in use-scenario tests. These objectives were accompanied by the ongoing goal of improving usability and our user interface. In addition, we added one significant new goal during the grant year: "storyboards." At the conclusion of the grant, all these goals had been met, with some partial exceptions, as follows:

• ***Data Mine and Import Metadata From SNAC*** *(original stated objective).* One of our aims was to work with the SNAC (Social Networks & Archival Contexts) project [described under "Environmental Scan" above] to data mine and import a subset of their metadata for our database. The purpose was to provide RoSE users with a larger foundation of pre-existing humanities bibliographical metadata in the system. Especially attractive to us was the ability to experiment with the particular kind of prosopographical metadata that SNAC harvests from finding aids.

For the original RoSE prototype, we had already stocked our database with extensive metadata about authors and works machine-gathered from Project Gutenberg and YAGO (the latter a semantic knowledge base generated from Wikipedia, WordNet, and GeoNames). Project Gutenberg provided us

with data for 11,964 people and 34,077 documents, but with no keywords or information about relationships. YAGO provided data for 7,557 people and 11,395 documents, with keywords offering opportunities for clustering studies and also "influence" identifiers to establish relations between people.

To initiate our new work with SNAC data we engaged with Daniel Pitti, Director of SNAC, and other SNAC lead developers, who generously provided data from their project. We worked with 125,000 individual SNAC XML files to harvest information useful for RoSE. The process began with parsing SNAC entries into selected component name, date, document, and keyword metadata and removing parsing errors to produce a clean, relatively compact set of XML files compatible with RoSE metadata format. Then, because the particular nature of SNAC's data sourced from finding aids produced a disproportionately large amount of material not well suited to RoSE (e.g., singleton items not connected to other items; many corporate authors or collecting entities), we undertook a second stage of processing. We identified principles for selecting well-connected items from the SNAC data that would help users see networks of people and works; and we also honed further methods for filtering, parsing, and cleaning up the information to match RoSE's mission. In doing so, we concentrated on the approximately 74% of the above-mentioned 125,000 SNAC files that are entries for people, and concentrated in addition on the 3% of these SNAC files representing highly connected people.

At the end of our grant, we had met our objective by selectively harvesting data from SNAC for 3,412 people and 28,036 works. The inclusion of SNAC enabled us to make interesting global scale comparisons between our three machine-harvested datasets. For example, in Figure 3 we compare author data we harvested from Yago, Project Gutenberg, and SNAC by historical birth year. (Although only a fraction of author data included birth dates--87% for Yago, 67% for Guttenberg, and 22% for SNAC--we project the actual number of authors per year in the data based on the ratio of those with dates to the total number of authors.) One finding we made is that these datasets tend to be complementary rather than redundant in coverage. Yago, based on Wikipedia, and Project Gutenberg, for example, overlap much less in authors than we originally expected, due primarily to the different periods of time they cover. By corollary, another finding we made was that YAGO, Project Gutenberg, and SNAC show a distribution of birth dates peaking around the years 1950, 1860, and 1900, respectively. Other characteristics of the collections such as the ratio of the number of unique surnames to the number of unique person entries seem to indicate a "familial preference" within some of the datasets, although this theory will have to be explored in more detail before any hard conclusions can be stated. Nonetheless, the statistical comparison of high-level inter-dataset features such as these
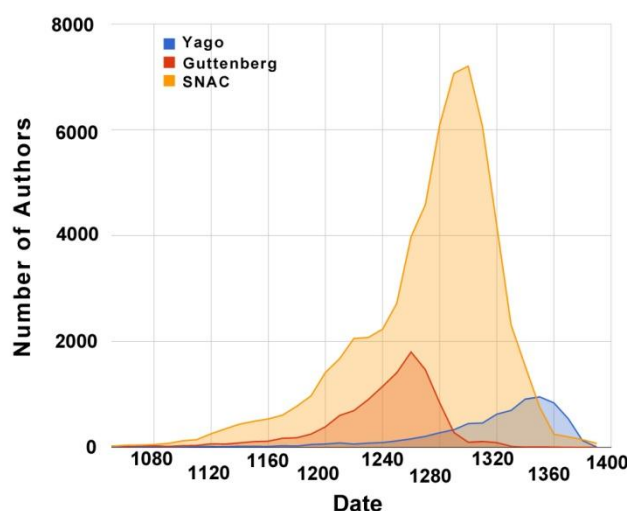

Figure 3: Graph of number of authors in our Yago, Project Gutenberg, and SNAC datasets by year. Our research shows that, as scholarly sources, these datasets are complementary rather than overlapping.

could reveal otherwise hidden patterns and be the basis for future data-mining research.

We also benefited in other ways from working with SNAC data. First, we were in effect trying out the paradigm by which one digital humanities project, in this case, RoSE, could act as a "client" of another digital humanities project, SNAC, where the client is positioned as a data-receptor for the other project's data source. This not only allowed us to make a suggestion to SNAC (that they consider creating what amounts to a selectable set of "channels" for particular datastreams--e.g., "humanities," "history," "literature," or "poetry" content-channels, as opposed to "science" channels)--but also encouraged us to think about how our own RoSE project might in the future serve client-projects of its own. RoSE currently has limited interoperability through XML/RDF export on an individual record basis. Secondly, we also benefited from comparing notes with the SNAC developers about data visualization. (The SNAC team began work on visualizations during our grant period.)

• *Improve Visualizations in RoSE; Add More Visualization Types* (original stated objectives). Another goal was to improve and expand RoSE's repertory of visualizations, which originally consisted of social-network and timeline graphs. Creating new, and improving existing, visualization types was challenging because our visualizations are dynamically generated; filterable; interactive (allowing a user to change their "point of view" in a network, for instance, by clicking on a different node); and information-rich (optionally reporting navigation history and other metadata in a sidebar). Also, we expanded our goals for visualization to support the important new "storyboard" function in RoSE (see below).
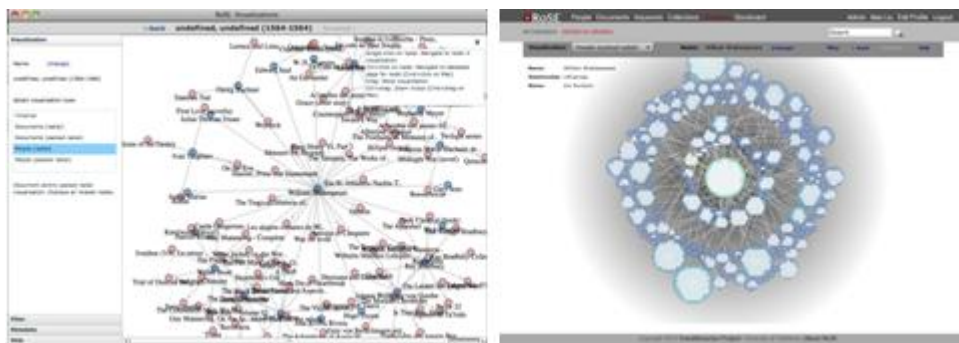
Figure 4: Social network and packed radial visualizations in RoSE

At the end of the grant period, we had met our objectives by improving our existing visualizations types (though the timeline graphs do not yet fully demonstrate the results we would like because of missing "date" information in many of our sources). We had also added (and continued to improve) a new radial "packed radial" visualization plus a means for users to produce, save, and export visual "storyboards." More detailed description of the design principles we employed in creating the RoSE visualizations may be found in **Appendix C**.

Finally, we undertook a major logical redesign of the algorithm generating visualizations from our database. To optimize real-time performance, we implemented a step-saving method of iteratively processing a whole "wavefront" of active nodes at a time (detailed explanation in **Appendix D**).
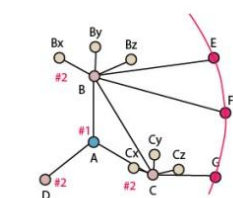

Figure 5: RoSE "wavefront" visualization algorithm

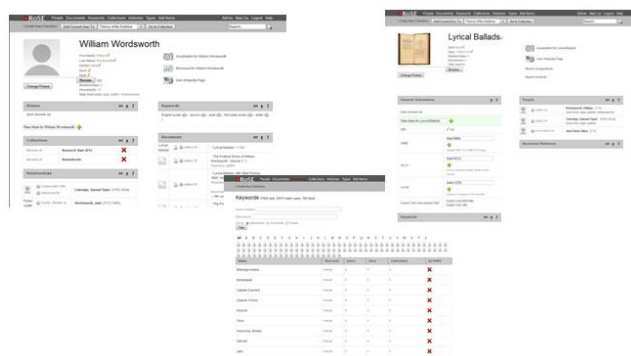• *Improve User Interface and General Usability (ongoing objective in addition to stated objectives).*

Figure 6: Examples of user interface screens in RoSE

During the year, we continuously and iteratively improved the RoSE user interface to simplify the presentation of data and search/edit functions; clarify features; improve the workflow by which users find, add, and edit data; integrate the database user interface with the visualization interface; and, of course, solve bugs and browser compatibility issues. To guide this work, we drew on feedback from our own developer group (who engaged in collective test sessions) and from students in our use-scenario study (see below).

However, we also knew we had to be realistic about how polished and smoothly functional our beta product could be, since otherwise the entire grant could have been spent on usability improvements to the detriment of other tasks. For this reason, we carefully prioritized usability issues and concentrated on those that were most important or relatively easy to fix, leaving additional improvements for a future implementation stage of the project.

At the end of the grant period, we had made significant progress on usability--especially in regard to issues that students identified in our use-scenario study. In addition, we created help documentation--including a Quick Start Guide, a demo video, and a "Learn More" suite of resources. However, a realistic assessment would be that we ended only at about 85% of the way toward "optimal" usability, where "optimal" means that a new user is unlikely to run at some point into a bottleneck (a confusing feature, a feature not yet fully debugged, or an inefficiency in work flow).

• *Create "Collections," "Histories," and "Storyboard" Features (new objective).* Our work on adding metadata to our database, creating visualizations, and improving usability led us midway through the grant year--as a logical extension of functionality improvements--to develop a way for users to collect and save their findings in RoSE. We thus created a "history"-tracking feature, which can be toggled on to automatically save a record of items traversed; and also a "collection" feature, which creates named, shareable collections of items (to which other users can "subscribe").
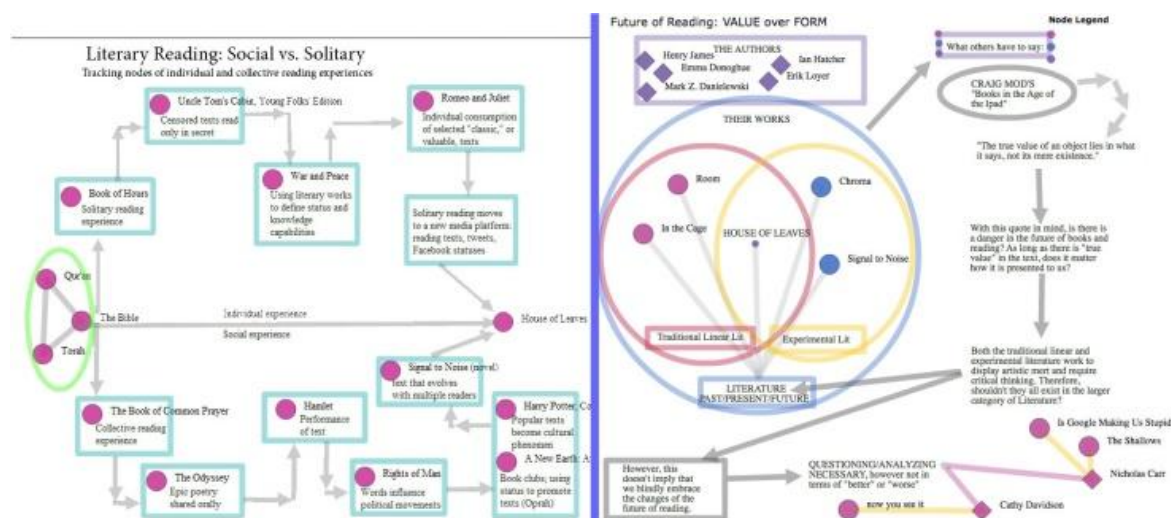
Figure 7: Examples of RoSE storyboards created by undergraduate students. (Courtesy of Kristin Crosier and Dani Williams.)

Then we had a breakthrough new idea: users should be able to work manually with their histories and collections so as to shape them into filtered, clarified, annotated, and otherwise interpreted proto-"arguments" or "narratives"--i.e., sketches of arguments about an intellectual topic that can be presented to others. An analogy might be the difference between communicating an idea as a bullet list of nouns and as a fully-formed, syntactical sentence. The logic of coordination and subordination required to form a sentence forces one to shape a loose set of ideas into an argument (or, as in a topic sentence, the beginnings of an argument).

We thus created an innovative "storyboard" feature that allows users automatically to populate a visual canvas with node-and-link representations of persons and works in their collections or histories. This is the equivalent in the analogy above of a bullet list. But then users can arrange, connect, color, annotate, and draw arrows and shapes around their data. This is the visual equivalent of writing a sentence. Finally, they can save their storyboard as an XML file and reload it later for revision (or load other people's storyboards). They can also export or print their storyboard as an image file (examples in Figure 7).

• *Run Use-Scenario Studies (original stated objective)*. Originally, we planned to study how RoSE operates in three real-life "use scenarios." Not the same as formal usability tests, these studies were appropriate to our prototype-to-beta development work because they could offer in-progress evaluation. Our intent was to identify opportunities and problems through observation of users and participant interviews and questionnaires.

The three use-scenario studies we planned included: (a) undergraduate classroom use, (b) professional conference or collaborative project use, and (c) individual-scholar use (e.g., researching a dissertation). This proved to be too

ambitious, however; in part because we also had to apply for human-subjects protocols to work with human test subjects (an aspect of research not traditionally familiar to humanities scholars). We thus decided to simplify by concentrating on the use-scenario study from which we expected the most important results: undergraduate classroom use. For use scenarios "b" and "c" above, we only performed informal studies.

A. Undergraduate Classroom Use-Scenario Study

We chose for our use-scenario study an undergraduate course co-taught at UC Santa Barbara in spring quarter 2012 by two of our project team: Rita Raley and Dana Solomon. The course (with 23 students) was English 146, "Distracted Reading," which explored the topic of "reading" practices with special attention to different media environments (https://engl146.wordpress.com/about/).   The instructors created one assignment for the course that required students to use RoSE. The assignment was described as follows (slightly abbreviated):

> Students will individually create what is called (1) a "collection" organized around a particular topic such as attention, distraction, and online reading, as well as (2) a "storyboard" that simplifies, narrates, or otherwise manually shapes their collection. These storyboards will visualize relations among authors and documents, such that each student will in effect be creating a network map that will both reflect the work we have done in the class (e.g. showing us how *House of Leaves* connects to Henry James) and produce new knowledge (e.g. how does your visualization show us something about the problem of distraction that we had not previously discussed). The objects you need for your collections may not all be in the RoSE system at present, so some of the work of this assignment may be data entry.
>
> We will demo the system in class and Dana Solomon will be holding regular lab hours in SH 2509, where you are

welcome to do the work for this assignment and for your final projects on computer-assisted reading.

Format of your work:
Storyboard analysis: On this page you will post a screen capture and short descriptive synopsis of your storyboard. When we are done, we will compare our different storyboards — our knowledge networks — in class and discuss the different readings we have produced of our topics. This discussion will be a prelude to the final course assignment, in which you will reflect more broadly on the transformations in humanities scholarship and pedagogy.

Reflections on using RoSE: On this page, you will also post a short reflective statement about your use of RoSE that addresses some of the questions posed above about the utility and value of the system for humanities scholarship. Some questions: What in your view are the possibilities and limitations of using a system such as RoSE in your coursework? How might you compare your storyboard with the traditional interpretive work you have done with some of these same texts and issues?

Use-scenario study: We will ask in class if you agree to have your comments used anonymously in the eventual reports that are prepared for the National Endowment for the Humanities, the granting agency supporting development work on RoSE. We will also ask if you agree to complete a questionnaire for the use-scenario study we are conducting. Your reflections on using RoSE will also be helpful for future development work, but you are not obligated to grant your consent.

Deadline: Friday, June 1 [we will discuss your storyboards along with the final assignment in class on June 6].

On May 21, 2012, our RoSE team ran the preliminary, in-class "demo" session mentioned in the assignment. Observation of this session, which we documented in a narrative report (**Appendix E**), showed a high level of interest and enthusiasm among the students, and also difficulties they had using the system. The flavor of the event is captured in the following excerpts from our observer's report:

As students began figuring out ways to add items and create collections, ejaculations of "cool" and "aha" began to circle. When a student figured out a function, he/she was quick to explain the move to surrounding peers. This collaboration between students to learn the system and work around areas of low functionality was especially encouraging, and generated a sense of camaraderie to keep exploring the RoSE site.

Students were especially excited by the prospect of adding new keywords and items. One . . . began adding a number of vintage advertisements from the 50s and 60s to the RoSE database. He appreciated the ability to add images of the advertisements to the document pages. . . .

Students had a lot of fun playing around with the "Relationships" feature. One student mischievously married Professor/RoSE supervisor Rita Raley to the novelist Henry

James. Other students called to friends across the room, "I'm your rival" and "I'm a scholar of your work."

An initial stumbling block was an error in the Collections functionality. . . . Another student . . . expressed some confusion over the different search box options. . . .

Despite some difficulties with the interface, many students saw potential for the nascent site. [One student] noted that the facts and information in the database were less interesting than the way users could explore how things fit together.

Following the demo session, students worked individually to fulfill the assignment. Materials they filled out or created in the process included: • a consent form; • an optional questionnaire for evaluating RoSE; • and their storyboards and blog-reflections on RoSE. (For the consent form and questionnaire form, see **Appendices F & G**.) Students posted their storyboards and reflections online at https://roseucsb.wordpress.com. (The RoSE site also includes a gallery of the storyboards used by permission: http://rosedocumentation.wordpress.com/2012/07/09/storyboard-gallery/).

After the assignment, we studied the results of the student questionnaires and also the students' reflections on RoSE (excerpts included in **Appendix H**: "Student Suggestions for RoSE Developers").[2] This provided us with crucial evaluative feedback, giving us both a sense of the problems the students had with the system and a sense that they recognized the system's potential. We were particularly impressed by the fact that, notwithstanding the then-extent usability problems, the students seemed to "get it" and were keen to discuss the pedagogic possibilities when they shared their storyboards in class during the final week of the term. We also learned that the students enjoyed being welcomed into the project in mid-development as participants whose feedback was valued.

After analyzing the student responses, we redesigned and reprioritized our development plan to address major suggestions and bug reports. Overall, our objectives for use-scenario studies were amply fulfilled in testing RoSE in this course.

B. Conference or Collaborative Use Scenario
In addition, we used RoSE informally to support and document a "conference or collaborative project use" (as originally stated in our grant proposal). The UC Santa Barbara English Department annually stages an end-of-year, capstone "Research Slam" event that allows graduate students, undergraduates, and faculty to show-and-tell their research projects in poster-show fashion. Many of these projects are in digital humanities or new media studies. In preparation for the department's fifth annual Research Slam

---

[2] Except where we have permissions from the students involved, student names are blacked out in this white paper and its appendices in compliance with our project's human subjects protocol.

on May 25, 2012 (http://lcm.english.ucsb.edu/?tag=research-slam), we used the "collection" feature in RoSE to prepare a research context for the event consisting of interrelated data about participants and the authors and works their projects addressed. (While we distributed feedback questionnaires for this use scenario, we received back only two completed questionnaires.)
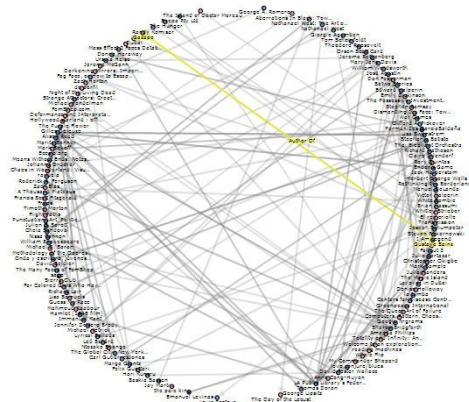

Figure 8: RoSE visualization of Research Slam "collection"

C. Individual-scholar Use Scenario

Finally, we also deployed RoSE informally for what our grant proposal called "individual-scholar use (e.g., researching a dissertation)."
One member of our project team, Dana Solomon, used RoSE to facilitate early research for his dissertation entitled "Reading Information Visualization in the Humanities." He created a collection of resources in RoSE related to the history of data visualization.
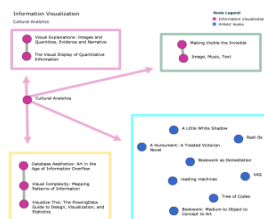

Figure 9: Storyboard visualization of History of Data Visualization

– –

In overview, we met our stated project objectives as well as additional objectives, though--as befits a beta project--there is more or less room for improvement in each area.

4. STATE OF PROJECT AT END OF GRANT PERIOD

RoSE Content and Feature List Summary

- An extensive database of bibliographical metadata (no full texts, but links to some texts and information elsewhere on the Web) for authors and works machine-harvested from the Project Gutenberg textbase, the YAGO knowledge base, and--with permission and collaboration--the SNAC (Social Networks & Archival Contexts) project

under development at the Institute for Advanced Technology in the Humanities at U. Virginia;

- An initial set of additional, user-entered bibliographical metadata and folksonomic-style "relationships" and "keywords" created by humanities scholars and students to thicken the pre-existing bibliographical metadata in relation to specific research topics;

- A user interface modeled as a social network site with "profile pages" for each author, work, and user in the system showing relations to other authors, works, and users;

- Editing functionality for adding and editing authors, works, relationships, and keywords;

- Dynamically-generated, interactive visualizations in several styles (social network graph, radial, packed radial, timeline) to facilitate seeing relationships in the system and navigating through the system;

- A "history"-tracking feature to create an automatic transcript of resources traversed in the system;

- A "collections" feature to harvest and share resources found in the system;

- A "storyboard" feature to populate a visual canvas with "collections" and "histories"; to manually filter, move, annotate, and draw shapes and arrows around these resources to present an argument or story; and finally to save the canvas as a shareable XML file;

- An XML export feature allowing metadata in the system to be exported on an item-by-item basis;

- User documentation, including "Quick Start Guide," "About RoSE," and "Learn More" (and also documentation for developers). (See http://rosedocumentation.wordpress.com/)

Technical Description

RoSE is a Web-accessible system developed in the open-source Ruby on Rails programming environment (http://rubyonrails.org/) on top of a MySQL relational database. It uses the open-source Adobe Flex framework (http://www.adobe.com/products/flex/) with the open source Flare library (http://flare.prefuse.org/) for visualizations. Information in RoSE is associated through keywords and user-customizable "relationship types" (e.g., "influence on," "collaborator of," "enemy of") that allow for the addition of thick description to the normally sparse social descriptors of standard metadata protocols (e.g., Dublin Core). RoSE is currently hosted on a UC Santa Barbara English Department Linux machine acting as a development server. (Previously, RoSE also ran on a second development server hosted by the

UC Santa Barbara Media Arts & Technology Program.) Backup of the system and its content occurs through routine backup of the English Department's servers (supplemented by a CD snapshot to capture the state of the project at the end of the grant period).

Project Dissemination and Publicity

While further work needs to be done in several areas, including usability improvements, metadata ontology, and what we call "authority transparency" (see "Reflections" below), RoSE is now at the stage where we are making it an open beta on a selective basis. Members of the public can explore RoSE as guest users by leaving the login fields on the home page blank and clicking "login." Scholars, teachers, students, and others with a special interest can request a codeword to register as an active user with a profile page and the ability to edit the knowledge base (contact for requests: ayliu@english.ucsb.edu).

The system is not yet at a stage where it can be open to all interested users. The main reasons are that the beta resides on a development server and lacks ongoing funding for "help" support, such as research assistants who can work with instructors to prepare for using RoSE in courses.

Project team members have begun presenting the project in research publications and talks. Existing and forthcoming research dissemination includes:

Published and Forthcoming Essays:

- **Liu, Alan**. "Friending the Past: The Sense of History and Social Computing," *New Literary History* 42.1 (2011): 1-30. Print.
- _____. "From Reading to Social Computing." *Literary Studies in the Digital Age: An Evolving Anthology*. Ed. Kenneth M. Price and Ray Siemens. MLA Commons, MLA, forthcoming. Web. http://dlsanthology.commons.mla.org/

Presented and Scheduled Lectures:

- **Liu, Alan**. "3 Key Digital Humanities Trends: How Digital Humanities Register Changes in the Humanities Today." National Humanities Center, 5 October 2012; and U. Pennsylvania, 12 November 2012.
- _____. "This is Not a Book: Transliteracies and Long Forms of Digital Attention." Translittératies Conference, École normale supérieure de Cachan, Paris. 7 November 2012.
- _____. "Remembering Networks: *Agrippa*, RoSE, and Network Archaeology." Miami U., 21 April 2012; U. Western Australia, 23 March 2012; U. of Guelph, 1 March 2011;; U. Amsterdam, 21 October

2011; Brown U., 3 October 2011; Carnegie Mellon U., 16 September 2010.
- _____. "RoSE (Research-oriented Social Environment)." Compatible Data Initiative meeting, New York City, 24 September 2011. (Talk presented via Skype.)
- _____. "RoSE (Research-oriented Social Environment).." U. California, Santa Barbara, 17 May 2010; 16 February 2010.
- **Liu, Alan, and Rama Hoetzlein**. "RoSE (Research-oriented Social Environment)." U. of Virginia, 19 April 2013.
- **Raley, Rita**. "RoSE (Research-oriented Social Environment)." New York U., 25 October 2012.
- **Solomon, Dana, and Lindsay Thomas**. "RoSE in Real Time: Social Science Research Methods in Digital Humanities Contexts." Panel on "Adapting Social Science Methods to Humanities Research." Modern Language Association convention, 6 January 2013.

Publicity (early publicity during development of beta):

- "UCSB English Professor Receives NEH Grant for Humanities Bibliographical Social Network" (press release and online video interview about RoSE). 11 January 2012. UCSB Office of Public Affairs Press. Published interview: http://www.ia.ucsb.edu/pa/display.aspx?pkey=2626 Online video: http://www.facebook.com/v/2963147679253

Future of Project

Currently, we anticipate the possibility of moving the project forward to an implementation and production phase through another stage of grant proposals and development (depending on the availability of the project leaders in light of other projects). Some of the main development tasks we think would need to be pursued in order to reach an implementation phase include:

- Adding "authority transparency" features (explained under "Reflections" below);
- Addressing additional usability issues;
- Replacing our Flash visualizations with HTML5 based ones written in JavaScript;
- Performing more cross-browser compatibility checking;
- Adding a repertory of scholarly and instructional examples of resources, collections, storyboards, etc. that could serve as "recipes" for use.

5. REFLECTIONS ON LARGER DIGITAL HUMANITIES ISSUES

In working on RoSE, we tried out specific solutions for implementing our hypotheses about the ultimate purpose of the system for the end user. But our work also necessarily engaged some of the larger, shared issues in digital humanities (DH) today. The following are our reflections on several of these issues, each representing both an opportunity and a problem for the field. Some reflections point forward to future solutions we would like to try. As we say in our epigraph, "what if . . . ?"

(A) Are the Dead Just Like Us? (Part I)

RoSE brings into convergence the "library" and "social network" models of knowledge so that historical authors and works are part of a single intellectual society together with their later scholars and students. The dead are "alive" in this society because their works, reputations, etc. change relationally as the whole of that society changes (e.g., when later authors or scholars find new meanings, priorities, and influences in the past, or edit linkages among persons and works to shape new understandings of movements). Shakespeare's profile page and network visualization in RoSE thus dynamically change depending on what else happens in the social-document-graph where he is a node.

In some profound way, this model is true to what contemporary humanists understand intellectual history to be: not a fixed repository of the past but a teeming, rich ecology of minds and works in which we can never be sure what intellectual gene lines presumed to be fixed or extinct are even now mutating to surprise the present.

But isn't there a limit to this model, at which point it becomes fiction? The limit is mortality, the brute fact of which (with its opposite, life) the humanities in particular acknowledge as part of their core subject matter--i.e., the lives, histories, times, tragedies, and comedies too of human being. The difference between the dead and the living in RoSE is this: Shakespeare (for example) may have a profile page, but a living user created it for him (or an algorithm created it through machine-harvesting from knowledge bases that, at the end of the chain of agency, were created by living bibliographers and scholars). In RoSE, therefore, Shakespeare's "living" agency is of a different order than that of registered users.

We wonder: how can humanists accommodate their deep commitment to the long past in a presentist social-network-site paradigm without willful fiction? For that matter, how will the social-network-site paradigm itself--too young to have seen an entire generation of users through life and death--age gracefully so as to hold the deep past in valued trust? Can the humanities, building their own variants of social network sites, contribute wisdom to that?

(B) Are the Dead Just Like Us? (Part II)

Humanities metadata--e.g., fields in a database or elements in an encoding scheme for author, type of work, date, etc.--can be incomplete and messy by comparison with well-behaved, fully-explicit bibliographic metadata in the contemporary sciences. Not only do the humanities typically track a smaller number of explicit bibliographical data fields than the sciences (e.g., we generally do not track "datasets" and "tools" of the sort cited in a system like ConceptVista), but problems with the fields we do follow grow exponentially the farther back we go in time (or, geo-culturally, the farther we span across societies and heritages). It's not just a matter of data erosion that can be measured quantitatively (fewer dates and names in the past, less information in general). It's a qualitative problem of the most fundamental kind: the problem of historical changes in the very "kinds" (or ontological, classificatory typologies) of data in relation to the overall ontology of their times.

The historical dead, and contemporary "other" people too, have different kinds of names, familial or occupational descriptors, date of birth reckoning systems (e.g., East Asian newborns aged "1" at birth), and so on. Many such metadata issues can be adjusted to fit in a contemporary database whose tables are designed for modern, Western ontologies (or at least left "null"). But there are metadata ontology issues imposed by the far past (and other cultures) that cannot be adjusted on a field-by-field basis because it is the whole relationality of the fields--the ontology system itself--that does not match. For example, the "author" field for Early Modern ballads (in many cases fillable only with "anonymous") occupies a different relational position in a publishing and circulation system in which it might have been the "publisher" or "bookseller" who was often the most author-like agent on the scene, commissioning and creating the idea for works (much as the recent mashup book *Pride and Prejudice and Zombies* from Quirk Press was really the idea of its publisher). Or, to take another example, what do we do with a "date of composition" of a poem from the past like this, "either May 2 to 6, 1790; or September 1792," where it is not just a matter of shoehorning the date into the database but of representing the fact that date-precision spans have different meanings in the past (e.g., in rural societies where there was no need for measures more exact than "midday" or "mid-spring").

Ramesh Srinivasan and his collaborators have suggested creating "flexible ontologies" negotiating and adapting between indigenous ontologies and standard ontologies (for example, so that the knowledge ordering system of a Native-American culture can be respected within a museum cataloguing or holding system). Such fluid ontology methods depend on ethnographic engagement with communities (whether or not facilitated computationally), which is to say on having access to living links to heritage.

Actual elders or other people must be consulted who can collaborate actively in adjusting the way things are named, classified, dated, etc.

We wonder how digital humanities projects can create or simulate active testimony of this sort for the deep historical past? What happens ontologically, epistemologically, and ethically when we substitute so-called "unsupervised" data-mining methods (e.g., "topic modeling' in DH) for active testimony?

(C) How Do We Trust Crowdsourced Knowledge?
A general issue in digital culture today is the evolving relation between expert knowledge (credentialed, organization-sanctioned, or top-down) and the new networked public knowledge (crowdsourced, aggregated, alternatively credentialed [e.g., becoming a Wikipedia administrator], or bottom-up). Aspects of this issue include: competing professional or epistemological standards (e.g., the difference between academic research and Wikipedia's "no original research" principle); different governance structures; and different standards of information authority, trust, or credibility.

This set of problems is now migrating into the humanities, with DH at the forefront. DH and related digital academic, information-study, and publication fields are importing into the world of academic expertise some of the ethos and practices of networked public knowledge--e.g., through crowdsourcing historical, translation, and other research projects. As a consequence, academic variants of the general issues attendant on crowdsourcing are appearing (in such exploratory permutations as collaborative research and publication, open peer review, open online or pre-print dissemination, open commentary, etc.).

RoSE is one such academic experiment in crowdsourcing knowledge. It is designed to crowd into the same database both expert / authoritative knowledge (information harvested from existing knowledge bases) and user-generated knowledge (user added or edited metadata). But currently RoSE has no method for reviewing, governing, or authorizing its crowd of knowledge (other than informal inspection by its developers). Thus nothing prevents a registered user from creating, for example, a person in the system called "Santa Claus" and then linking that person as "author of" to a fictitious or real book. Nor does anything prevent a more well-intentioned user from entering factually inaccurate information.

We have an idea, though, for an approach to this problem that we intended to try, but could not fit into our NEH grant year on top of more immediate goals. We call our idea "authority transparency," which is based on earlier ideas by Hoetzlein (2009) and Liu (2007). In the case of RoSE, the essential notion is that before registered users can edit or create information, they should be required to designate a named "project" (which can be a shared project) with which the actions they are about to take will be associated. For example, a project could be a dissertation, an undergraduate essay assignment, a shared class assignment, a conference, etc. In addition, all registered users would be classified by user type. In the context of our primary academic audience, such types might be "professor," "teacher," "graduate student," "undergraduate," "high-school student," etc. Then we would build into RoSE a mechanism for identifying and filtering search results in our database by user type and project type (or by individual project).

Our hypothesis is that transparency by user- and project-type amounts to transparency by--and customization of-- "authority," meaning that the standard of desired authority is set by end-users for their particular purpose. Another way to describe "transparent authority" is thus "appropriate authority." For example, in many cases of casual use, no filters need be set for user or project type. For others, a researcher might well want to concentrate only on material harvested by machine or entered by professors. Alternatively, for still other purposes, a researcher might want to see only information created by non-professors (e.g., while conducting research on student learning or the undergraduate understanding of a topic).

(D) "Building" as Understanding
One of the controversies now vexing the DH field concerns the question, "Do you have to be a builder [or coder]?" (Ramsay). Closely related is the debate about "hack vs. yack" (see, e.g., Cecire; Koh). The controversy, in other words, is about the proper balance between programming work and interpretive or theoretical work in DH. To be a digital humanist, does one have to hack vs. yack (or, a fair requital to builders, hack *and* yack)?

Many on our RoSE project team have a history of working on collaborative digital projects, including the earlier RoSE prototype designed in the Transliteracies initiative. In addition, RoSE co-director and Project Scientist Rama Hoetzlein brought the benefit of his experience on a collaborative NSF IGERT research team into both the Transliteracies and NEH-grant stages of RoSE. He also co-authored an article on the collaborative process involved in the Transliteracies stage of the project (Chuk, Hoetzlein, Kim, and Panko, "Creating Socially Networked Knowledge through Interdisciplinary Collaboration").

As a result, our RoSE team sees a more constructive way to frame the "do you have to be a builder?" issue.

First, our experience indicates that in today's increasingly team-oriented projects requiring multiple expertises, it is increasingly obsolete to inquire whether "one" is a builder or interpreter. The operative unit is the team, which, when structured around best practices of collaboration, performs both programming and interpretive/theoretical work in

tandem and, as a bonus, involves individuals with strengths in one area in learning about other areas (e.g., programmers learning about humanities author-attribution issues; or humanities scholars learning about the underlying principles of database, visualization, and user-interface design). Some of the best practices for collaboration we followed included: involving all hands in the conceptual design of the project; ensuring that there was individual research interest to be gained by each participant; and scheduling work so that programming, interpretation, and content creation do not wait on each other in a way that makes one the "initiator."

Secondly, we came to the conclusion that the question of whether DH developers are programmers and / or interpreters is in the last analysis misplaced. The proper locus of the question should be the end user, meaning that the most constructive question is whether DH developers can work together in ways that not only span across the programming / humanities interpretation divide in their own activities but also design this whole span of activity into their product itself as affordances for the user. It is the user who should be encouraged to be both a "builder" and "interpreter."

The "storyboard" feature we created in RoSE was our wake-up call in this regard. We were surprised how clearly the purpose and shape of our project came into focus for ourselves once we decided that the end-user should get out of our system the ability to "build" a visual canvas "interpreting" an argument. We were surprised in equal measure to witness the effect the storyboard had on undergraduate students in our use-scenario study. Project team member Dana Solomon reports:

> I was surprised at the degree to which the students in the "Distracted Reading" course took to the Storyboard feature. I had initially expected the students to be more interested in the "readymade" or "one-click" visualizations, due in large part to the fact that they are easier to access and essentially produce a finished product. However, the storyboard feature presents an opportunity for the students to produce a kind of visual narrative of their research, argument, collection, etc. Though the storyboard can ultimately be used to produce a static image, it harbors the fruitful possibility of producing an argument that unfolds diachronically in space and time. The students, and other users I've discussed the project with, seem also to enjoy the relative simplicity of the storyboard, as compared to the sliders, filters, and other technical functions available in the Visualization tool. In the broader context of visualization in the digital humanities, there is a great deal of excitement over powerful visualization engines used in projects like the Software Studies Initiative's Cultural Analytics series and Stanford's experiments with the Gephi platform. These projects present opportunities to render extremely detailed and advanced visualizations of data. However, these projects also call for a higher general degree of technical literacy and, I would argue, a higher level of critical literacy with respect to back-end processes generally effaced by the visualization interface. RoSE's Storyboard is emblematic of a different approach to visualization; it is as

minimalist and intuitive as something like a tag-cloud generator, but is conceptually much richer because it prompts construction, experimentation, and an almost tactile interaction that is absent in "one-click" visualization tools.

The RoSE Storyboard seems well-suited to a variety of humanities contexts. Based on my experience in the classroom, the RoSE Storyboard is especially useful for assignments that ask students to produce and perform knowledge or research, rather than simply locate a specific fact, connection, reference, etc. The Storyboard is always already self-reflexive to some degree in that it foregrounds the way we organize information and construct arguments. Assignments based around the storyboard therefore function on two levels: 1.) they engage student thinking in a particular way that revolves around the active production of or interaction with information/knowledge, rather than the pursuit of static information and 2.) these assignments work on a meta-level to recover the mental process of organizing information into narrative, argument, "story," or knowledge. The Storyboard function is also beneficial for researchers in that it provides an "informal," yet powerful, napkin-drawing style of information organization. The tool provides a chance to "sketch" out a rough idea of how different concepts might fit together, without knowing exactly where one might end up. This might be termed the "exploratory" function of the storyboard. Finally, the Storyboard might offer more creative practitioners a new medium for critical performance à la the kind of live-programming/live-coding I have seen in the past at various digital-themed conferences.

In essence, the RoSE storyboard allows the end-user to be a visual programmer of knowledge. Hack meets yack.

– –

Many other unsolved or complex issues in RoSE could be mentioned as well--e.g., the treatment of multiple versions and editions of works; the appropriate level of granularity with which an author or work is treated (was it 'William Wordsworth" who was influential, for example, or mainly "the early Wordsworth" and some poems, or even just some lines, in *Lyrical Ballads*?); and scale and scope problems (how many nodes to show at a time in a graph?).

But the issues we reflect on above serve as a demonstration of the payoff of a beta project like RoSE prior to any possible future implementation. A project of making and interpreting, RoSE is a digital humanities exploration of a new paradigm for the humanities. It explores the opportunities and problems--technical and theoretical--of friending the humanities knowledge base to the paradigm of the social network.

– –

RESEARCH LITERATURE CITED

[1] **Bérubé, Michael, Hester Blum, Christopher Castiglia, and Julia Spicher Kasdorf**. "Community Reading and Social Imagination." *PMLA* 125.2 (2010): 418.25. Print.

[2] **Cecire, Natalia**, "Introduction: Theory and the Virtues of Digital Humanities." *Journal of Digital Humanities* 1.1 (Winter 2011): 44-53. Web (PDF version). http://journalofdigitalhumanities.org/files/jdh_1_1.pdf

[3] _____. "When Digital Humanities Was in Vogue." *Journal of Digital Humanities* 1.1 (Winter 2011): 54-58. Web (PDF version). http://journalofdigitalhumanities.org/files/jdh_1_1.pdf

[4] **Chuk, Eric, Rama Hoetzlein, David Kim, and Julia Panko**. "Creating Socially Networked Knowledge through Interdisciplinary Collaboration." *Arts and Humanities in Higher Education* 11.1-2 (2012): 93-108. Print.

[5] **English, James F**. "Everywhere and Nowhere: The Sociology of Literature After "the Sociology of Literature." *New Literary History* 41.2 (2010): v-xxiii. Print.

[6] **Geertz, Clifford**. *The Interpretation of Cultures: Selected Essays.* New York: Basic, 1973. Print.

[7] **Hoetzlein, Rama**. "Alternatives to Author-centric Knowledge Organization." Implementing New Knowledge Environments conference (INKE 2009). Victoria, Canada. 23 Oct. 2009.

[8] **Koh, Adeline**. "More Hack, Less Yack?: Modularity, Theory and Habitus in the Digital Humanities." Adeline Koh, 21 May 2012. Web. http://www.adelinekoh.org/blog/2012/05/21/more-hack-less-yack-modularity-theory-and-habitus-in-the-digital-humanities/

[9] **Liu, Alan**. "Network Knowledge: Policing Web 2.0." New Network Theory conference. University of Amsterdam, 28 June 2007. [Later presented at various venues under the title "Peopling the Police: A Social Computing Approach to Information Authority in the Age of Web 2.0."]

[10] **McKenzie, D. F**. *Bibliography and the Sociology of Texts*. Cambridge: Cambridge UP, 1999. Print.

[11] ***Pride and Prejudice and Zombies: The Classic Regency Romance-- Now with Ultraviolent Zombie Mayhem***. By Seth Grahame-Smith and Jane Austen. Philadelphia: Quirk, 2009. Print.

[12] **Ramsay, Stephen**. "On Building." Stephen Ramsay, 11 January 2011. Web. http://stephenramsay.us/text/2011/01/11/on-building.html

[13] **Srinivasan, Ramesh, Alberto Pepe, and Marko Rodriguez**. "A Clustering-based Semi-automated Technique to Build Cultural Ontologies." *Journal of the American Society of Information Science and Technology* 60.3 (2009): 608-20. Print.

[14] **Srinivasan, Ramesh, and Jeffrey Huang**. "Fluid Ontologies for Digital Museums." International Journal on Digital Libraries 5.3 (2005): 193-204. Print.

PROJECTS CITED

[1] **Academia.edu**. 2012. Academia.edu. 2012. Web. http://academia.edu

[2] **Collex**. Version 1.6.7. N.d. NINES.org. Web. http://www.nines.org/

[3] **CommentPress**, Version 3.3.2. 2012. Institute for the Future of the Book. Web. http://www.futureofthebook.org/commentpress/

[4] **ConceptVista**. N.d. GeoVista Center, Pennsylvania State U. Web. http://www.geovista.psu.edu/ConceptVISTA/

[5] **Crowded Page**. N.d. Center for Digital Research in the Humanities, U. of Nebraska-Lincoln. Web. http://www.crowdedpage.org/

[6] **Goodreads.** 2012. Goodreads Inc. Web. http://www.goodreads.com/

[7] **Knalij for PubMed**. 2012. iWakari. Web. http://knalij.com/

[8] **LibraryThing**. N.d. Web. http://www.librarything.com

[9] **Living Social (Books).** N.d. Facebook. Web. [Applications closed down by Facebook on 12 Aug. 2011].

[10] **Mapping the Republic of Letters**. N.d. Stanford University. Web. https://republicofletters.stanford.edu/

[11] **MediaCommons**. N. d. Institute for the Future of the Book. Web. http://mediacommons.futureofthebook.org/

[12] **Open Journal Systems.** N.d. Public Knowledge Project. Web. http://pkp.sfu.ca/?q=ojs

[13] **The Orlando Project**. 2010. U. Alberta. Web. http://www.arts.ualberta.ca/orlando/

[14] **PreE (Professional Reading Environment)**. N.d. Electronic Textual Cultures Laboratory, U. Victoria. [Unreleased development project]

[15] **Project Gutenberg**. 27 Oct. 2012. Web. http://www.gutenberg.org/

[16] **Quanta**. By Rama Hoetzlein. Developed as part of his "The Organization of Human Knowledge: Systems for Interdisciplinary Research." MA thesis. University of California, Santa Barbara, 2007. Demo video: http://www.rchoetzlein.com/quanta/

[17] **Scratch.** N.d. Lifelong Kindergarten Group, MIT Media Lab. Web. http://scratch.mit.edu/

[18] **SEASR**. N.d. SEASR.org. Web. http://seasr.org/

[19] **SNAC (Social Networks and Archival Context).** N.d. Institute for Advanced Technology in the Humanities, University of Virginia. http://socialarchive.iath.virginia.edu/

[20] **TAPoR**. Version 2.0. TAPoR Project. 2012. Web. http://www.tapor.ca/

[21] **Transliteracies.** N.d. University of California. Web. http://transliteracies.english.ucsb.edu/

[22] **Visual Bookshelf**. N.d. Facebook. Web. [Application closed down by Facebook on 12 Aug. 2011].

[23] **Yaddo Circles Prototype**. See Micki McGee, "Yaddo Circles Prototype: Relationship Mapping Demo." N.d. Vimeo. Web. http://vimeo.com/36929545

[24] **Yago**. 3 Dec. 2012. Max Planck Institut Informatik. Web. http://www.mpi-inf.mpg.de/yago-naga/yago/

[25] **WorldCat Identities.** 2010. OCLC Online Computer Library Center, Inc. Web. http://www.worldcat.org/identities/

[26] **Women Writers Project: JIT hypergraph networks**. See John Melson and Julia Flanders, " Not Just One of Your Holiday Games: Names and Name Encoding in the Women Writers Project Textbase." White Paper to the NEH Office of Digital Humanities for Level 2 Digital Humanities Start-Up Grant. January 2010. National Endowment for the Humanities. Web. https://securegrants.neh.gov/PublicQuery/main.aspx?f=1&gn=HD-50320-08

AUTHOR BIOGRAPHIES

Ivana Andjelkovic is a Ph.D. candidate in the Media Arts and Technology Program at UC Santa Barbara. Her dissertation research focuses on audio analysis, music recommendation, discovery and visualization. She has a keen interest in music perception and cognition. She holds a Bachelor's degree from the University of Belgrade, Serbia, and a Master's degree from George Washington U., both in Computer Science.

Salman Bakht is a new media artist and composer who completed his Ph.D. in 2011 in the UC Santa Barbara Media Arts and Technology Program. His dissertation, was titled "Circumaurality (Listening around Sound): Representing Spatial and Media Contexts in Sonic Art." Salman's work focuses on the reproduction and transformation of recorded audio through algorithmic and natural processes and the analysis, representation, and integration of the physical environment and the media landscape.

Joshua Dickinson is a M.A. candidate in the Media Arts and Technology Program at UC Santa Barbara. He is a composer and computational artist whose research involves human-computer interaction, cultural analytics, and artificial creativity. He holds a B.A. in music composition from Columbia U., and a M.S. in multimedia engineering from UC Santa Barbara.

Michael Hetrick is a Ph.D. student in the Media Arts and Technology department at UC Santa Barbara. A lifelong musician, he discovered his passion for electronic music while studying at Western Reserve Academy. At Vanderbilt University, he expanded his work into the field of video art while receiving a B.A. in Digital Media and Distribution. He went on to receive an M.A. in Electronic Music and Sound Design at UC Santa Barbara in 2011 while doing research in chaotic synthesis under Curtis Roads. Today he is the co-owner and co-founder of Unfiltered Audio, a company dedicated to creating new software for all digital musicians.

Rama Hoetzlein is a researcher and scientist in knowledge systems and databases. He completed a B.F.A. in Fine Arts and a B.A. in Computer Science at Cornell in 2001. For his masters in the Media Arts and Technology program at UC Santa Barbara, in 2007, he developed a novel semantic database, Quanta, to explore representations of interdisciplinary knowledge. From 2007 to 2010, his dissertation work focused on issues of creative freedom in the media arts. In 2011 he was Professor of Media Studies in the Department of Architecture and Media Technology (Medialogy) at Aalborg University, Copenhagen. His current research focuses on knowledge systems and intelligent agents.

Andrew Kalaidjian is a Ph.D. candidate in English at UC Santa Barbara. His dissertation, "Places of Rest," is a study of environmental aesthetics in the modernist novel. His research interests include literature and environment, history of science, and the digital humanities.

Alan Liu is Professor of English (and affiliated faculty of Media Arts and Technology) at UC Santa Barbara. He has published books titled *Wordsworth: The Sense of History* (1989); *The Laws of Cool: Knowledge Work and the Culture of Information* (2004); and *Local Transcendence: Essays on Postmodern Historicism and the Database* (2008). Recent essays include "Friending the Past: The Sense of History and Social Computing" (2011), "Where is Cultural Criticism in the Digital Humanities?" (2012) and "The State of the Digital Humanities: A Report and a Critique" (2012). Projects he has directed include the University of California Transliteracies Multi-campus Research Group (2005-10) on online reading. Liu is co-founder and co-leader of the 4Humanities advocacy initiative.

Eric Nebeker is Acting Co-director of the Early English Broadside Ballad Archive and Lecturer in English at UC Santa Barbara. He recently completed his Ph.D. at UC Santa Barbara with a dissertation on "The Broadside Ballad and English Literary History, 1540-1700." His research articles on broadside ballads and English literary history have appeared in *ELH* and *Studies in English Literature*.

Rita Raley is Associate Professor of English at UC Santa Barbara. She has authored one book, *Tactical Media* (2009), and co-edited the *Electronic Literature Collection* (2011). Recent essays include "TXTual Practice" (forthcoming, 2013), "Dataveillance and Countervailance" (2012), "Another Kind of Global English" (2012), and "'Living Letterforms': The Ecological Turn in Contemporary Digital Poetics" (2011).

Dana Solomon is a Ph.D. candidate in English at UC Santa Barbara. He has presented papers on information visualization and electronic and contemporary literature. He is currently writing a dissertation entitled "Reading Information Visualization in the Humanities." Solomon is a HASTAC scholar and a member of the 4Humanities advocacy initiative, and is currently serving as the Arnhold Graduate fellow for undergraduate research at UC Santa Barbara.

Lindsay Thomas is a Ph.D. candidate in English at UC Santa Barbara. Her dissertation, entitled "State of the Unknown: Speculation and Security in Contemporary Media," concerns the relationship between security and speculation in contemporary speculative fiction and in governmental, international, and corporate computational media and technologies. She also works as lead research assistant for 4Humanities, an international humanities advocacy collective.

APPENDICES FOLLOW BELOW

APPENDICES

APPENDIX A
Sample Screenshots from RoSE

Figure 10: RoSE Home Page (http://rose.english.ucsb.edu/).
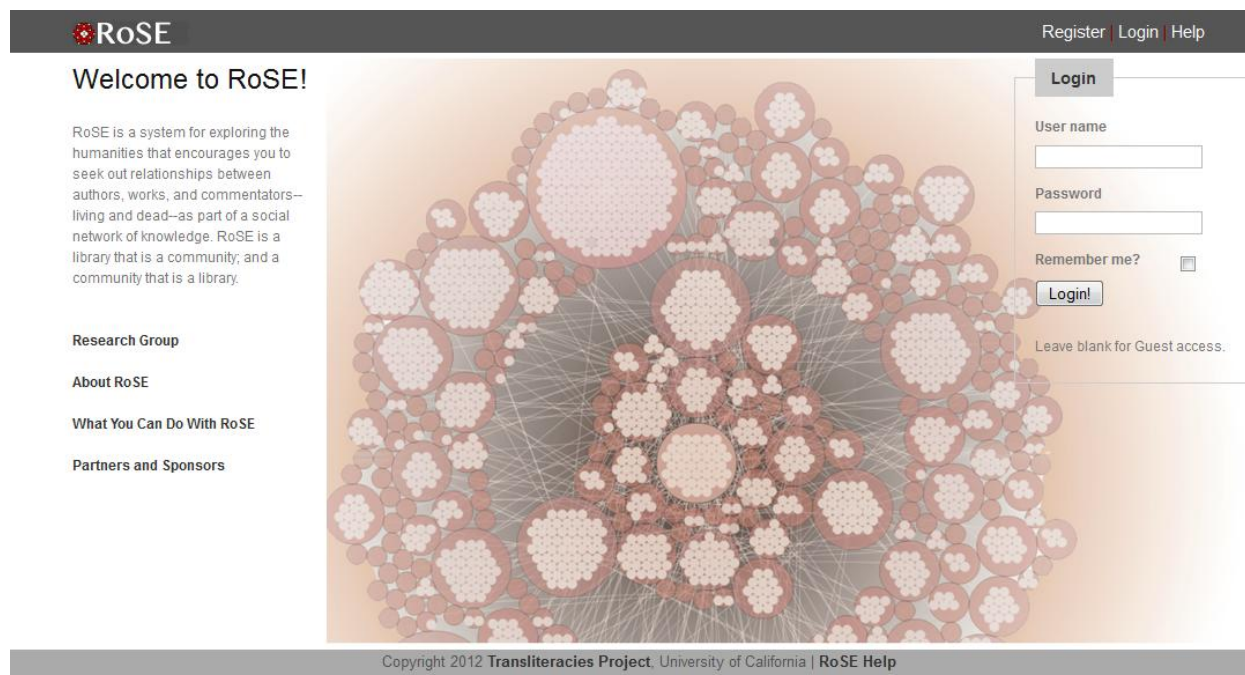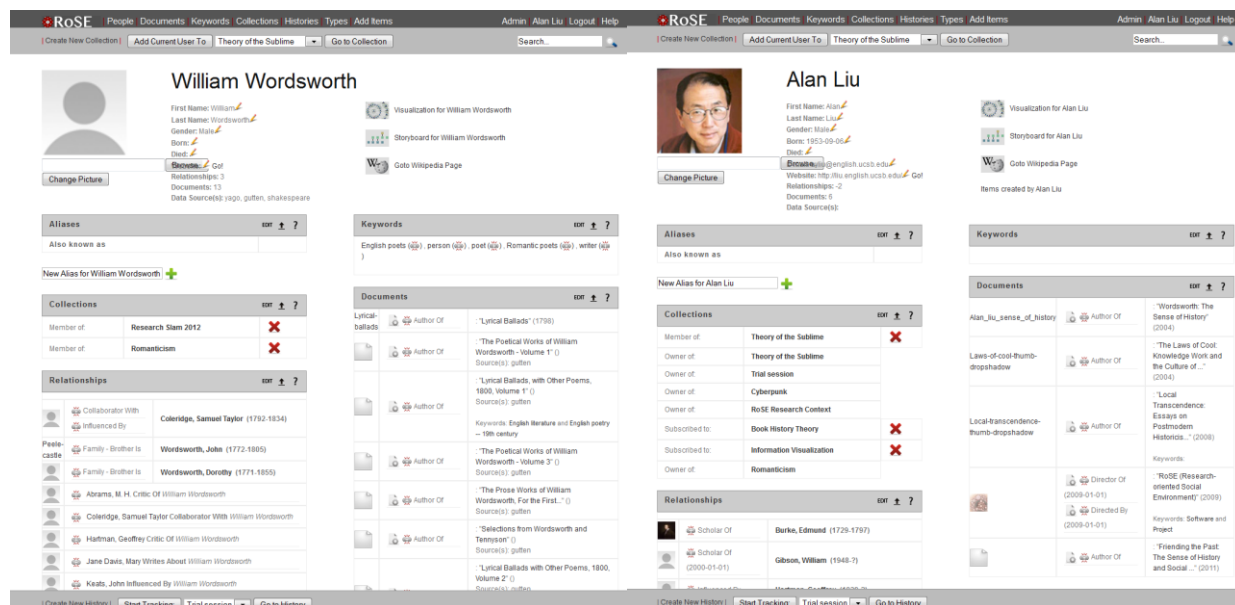
Figure 11: Profile pages for historical person (Wordsworth) and registered user (Liu), with relationships linking the latter to the former as "scholar of."
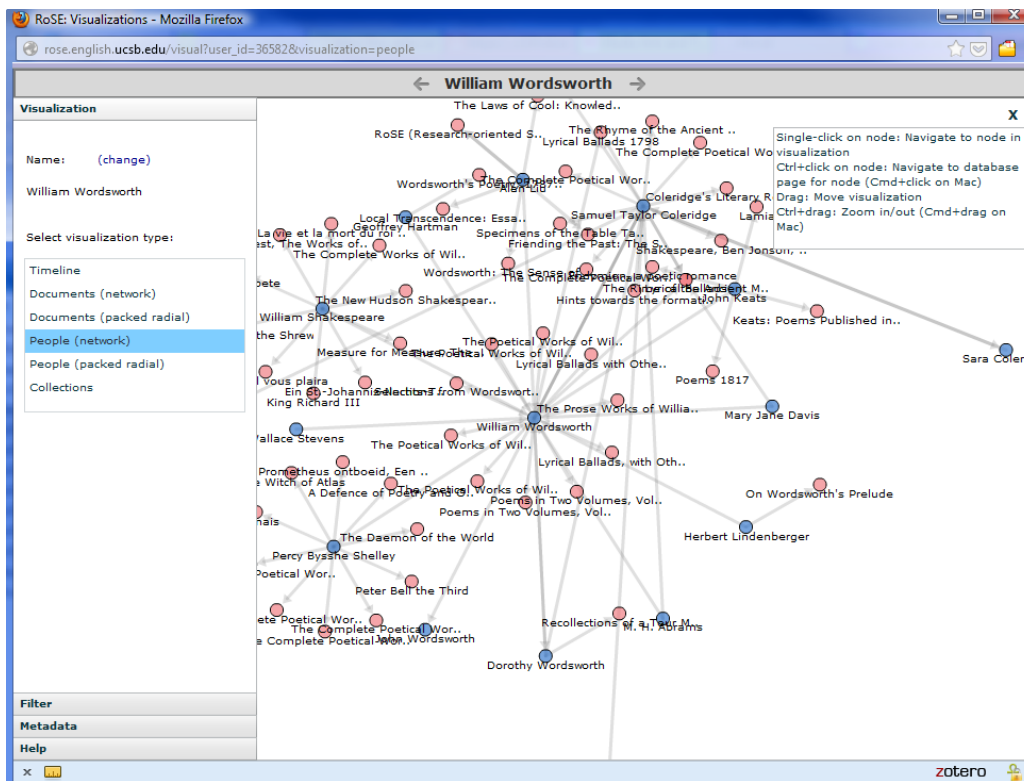
Figure 12: Social network graph visualization of author.
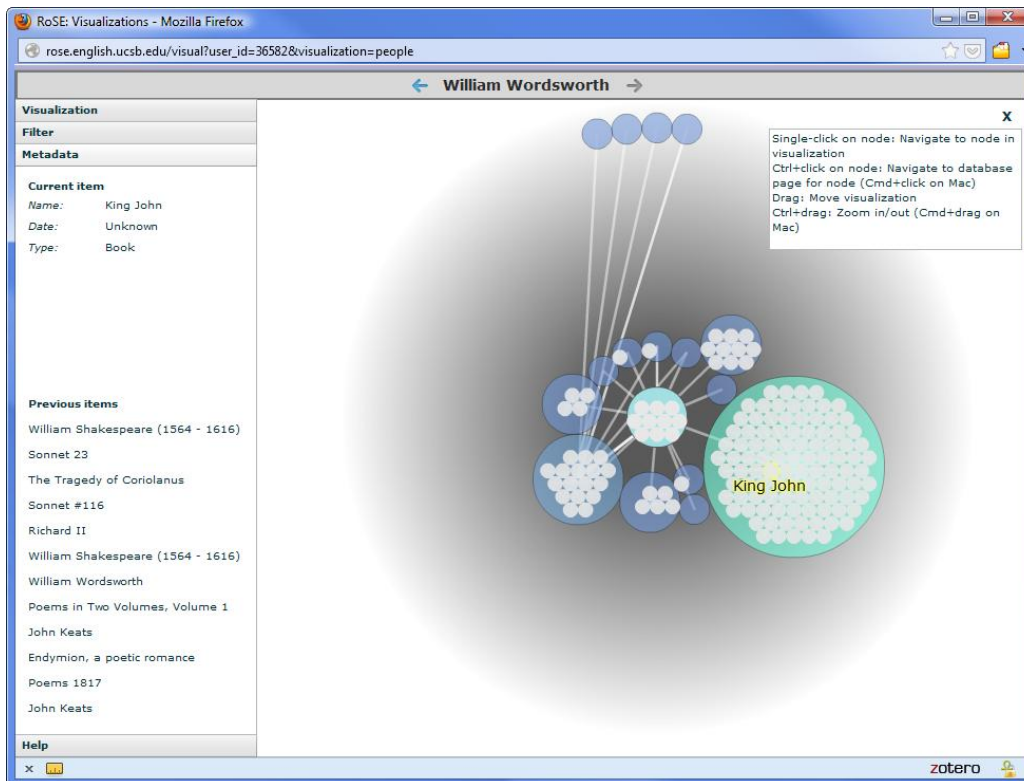
Figure 13: Packed radial visualization of author and related authors / works (with metadata in sidebar).
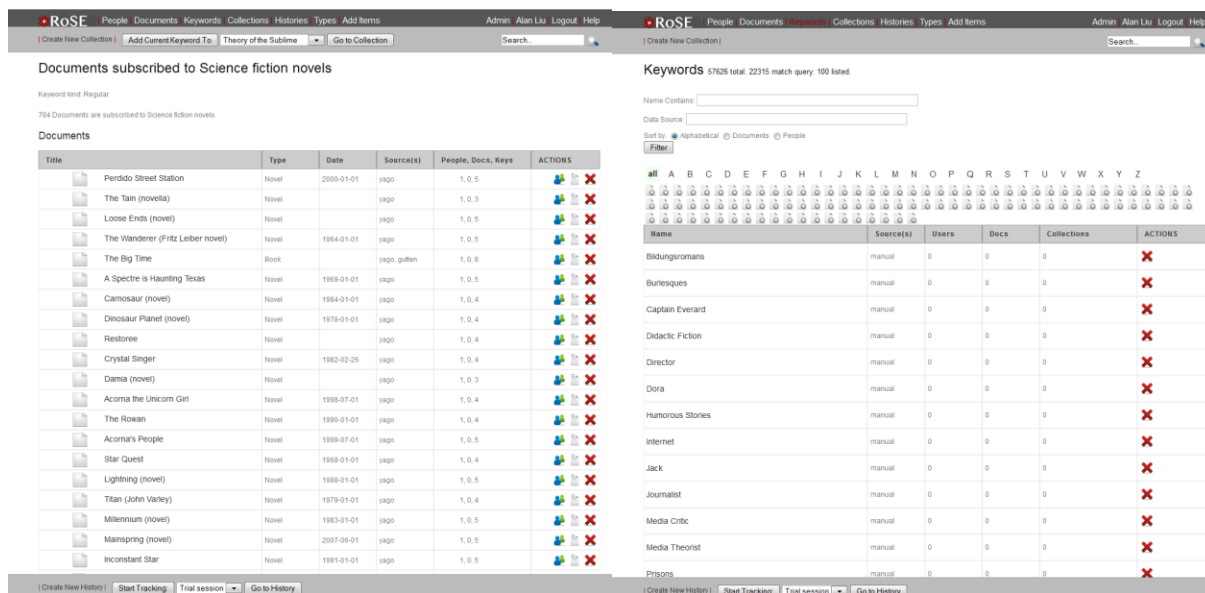
Figure 14: Document and Keyword lists.
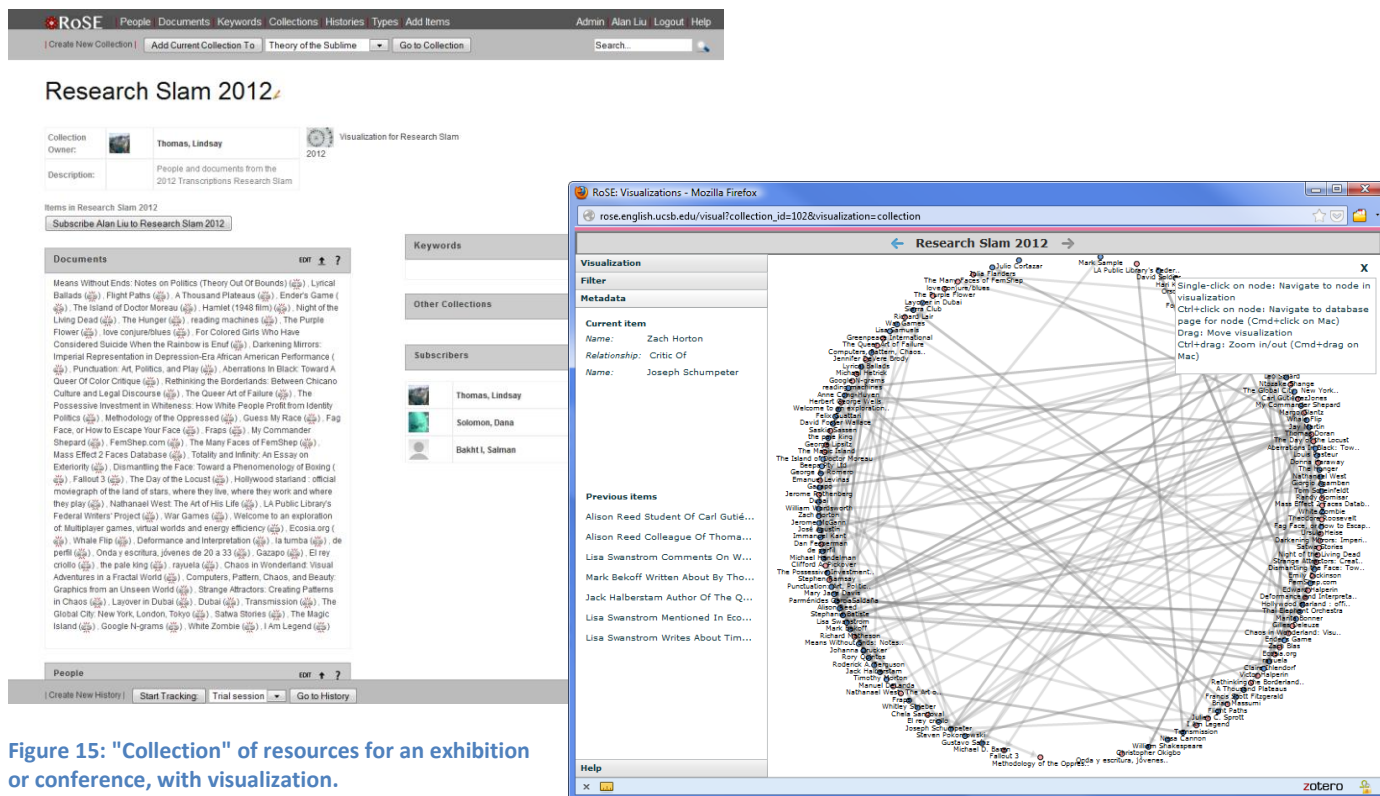
Figure 15: "Collection" of resources for an exhibition or conference, with visualization.

Literary Reading: Social vs. Solitary
Tracking nodes of individual and collective reading experiences

Uncle Tom's Cabin, Young Folks' Edition
Censored texts read only in secret

Romeo and Juliet
Individual consumption of selected "classic," or valuable, texts

Book of Hours
Solitary reading experience

War and Peace
Using literary works to define status and knowledge capabilities

Solitary reading moves to a new media platform: reading texts, tweets, Facebook statuses

Qur'an

The Bible — Individual experience / Social experience — House of Leaves

Torah

Signal to Noise (novel)
text that evolves with multiple readers

The Book of Common Prayer
Collective reading experience

Hamlet
Performance of text

Harry Potter, Co...
Popular texts become cultural phenomen

A New Earth: A...

The Odyssey
Epic poetry shared orally

Rights of Man
Words influence political movements

Book clubs; using status to promote texts (Oprah)

Figure 16: Examples of "storyboards" created by students.

Future of Reading: VALUE over FORM

Node Legend

THE AUTHORS
Henry James
Emma Donoghue
Mark Z. Danielewski
Ian Hatcher
Erik Loyer

What others have to say:

CRAIG MOD'S "Books in the Age of the Ipad"

THEIR WORKS

Room

Chroma

In the Cage

HOUSE OF LEAVES

Signal to Noise

"The true value of an object lies in what it says, not its mere existence."

With this quote in mind, is there is a danger in the future of books and reading? As long as there is "true value" in the text, does it matter how it is presented to us?

Traditional Linear Lit

Experimental Lit

Both the traditional linear and experimental literature work to display artistic mert and require critical thinking. Therefore, shouldn't they all exist in the larger category of Literature?

LITERATURE PAST/PRESENT/FUTURE

Is Google Making Us Stupid?

The Shallows

Nicholas Carr

However, this doesn't imply that we blindly embrace the changes of the future of reading.

QUESTIONING/ANALYZING NECESSARY, however not in terms of "better" or "worse"

now you see it

Cathy Davidson

Figure 17: RoSE user documentation and help pages (http://rosedocumentation.wordpress.com/).
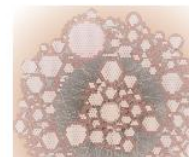
Figure 18: RoSE demo video
(http://rosedocumentation.wordpress.com/2012/11/16/demo-video/)

<u>A</u>ppendix B
Selected Transliteracies Research Reports
and Papers Relevant to RoSE

The following is a selection of the topics of 78 research reports and 7 research papers produced for the Transliteracies initiative as the initial environmental scan for RoSE. Full texts of the reports and papers may be found at http://transliteracies.english.ucsb.edu/category/research-project/research-clearinghouse. (We have also produced a larger number of briefer related reports under the heading "Objects for Study" at the above location.)

(1) Reports on advanced humanities online document reading, bibliographical, and research environments:
- Collex
- CommentPress
- Document Database Integration for the Professional Social Environment (including PreE/REKn)
- Electronic Book Review (EBR)
- MediaCommons
- Open Journal Systems
- Sophie
- Zotero

(2) Reports on social book cataloging sites and/or social academic networks:
- Academia.edu
- Freebase
- LibraryThing
- Social Book Cataloging: Humanizing Databases

(3) Reports on closest analogues to RoSE:
- ConceptVista
- SNAC (Social Networks and Archival Context)
- WorldCat Identities

(4) Other reports on projects, tools, and initiatives relevant to the following aspects of RoSE:

- Reports on metadata issues:
  - o Narrative as Metadata

- Reports on social-network analysis tools (and text-mining tools useful for studying social computing):
  - o Blogdex
  - o ConceptNet

- Reports on visualization methods and tools:
  - o A Comparison of Development Platforms for Social Network Data Visualizations
  - o Visualization Ecologies
  - o TextArc

- Reports on advanced humanities text-/data-mining and pattern-recognition:
  - o MONK and SEASR.

APPENDIX C
Design principles for RoSE visualizations

Some of the questions that guided design of visualizations are: (1) How to display large amount of data in such a way that it facilitates navigation and knowledge discovery? (2) How to represent key figures, documents and development over time in an area of interest? (e.g. Modernism, Science Fiction), (3) How to enable users to organize their findings, create coherent story and share it with others? Accordingly, three basic types of visualizations have been created in order to provide users with unique views of the data:

1. Temporal view – "Timeline" visualization shows people, documents and their relationships arranged linearly based on associated dates.
2. User/document centric view – "Network" and "Packed radial" visualizations focus on a particular person or document and direct and indirect relationships to other items in the database.
3. Storytelling view – "Storyboard" allows user to visualize narrative by displaying collections, annotating items, drawing and modifying the appearance of nodes.

Some characteristics of each type are described below, in order to demonstrate solutions to encountered design challenges.

Users and documents – 2 entities in one visualization

Unlike most social network visualizations that focus on relationships between people, RoSE visualizations deal with two different data types - people and documents. Both types are of equal importance for RoSE users and the visualizations treat them as such. Hence, we wanted to distinguish between groups of documents and groups of people in all visualizations, and represent tight interrelationships at the same time. In addition to color coding of people and documents, we designed layouts accordingly. In the Timeline visualization, we display both groups in two separate, yet connected, streams. In user and document centric visualizations (Network and Packed radial), we group users and documents based on their relationships, while creating views focused on people or documents (People network and packed radial, Documents network and packed radial).

Navigation through large amount of data

Given the large amount of data and potentially high interconnectedness, we chose to limit the amount of data that can be visualized. For example, in Timeline visualizations, we display only data tagged with a particular keyword. In user and document centric visualizations, we display nodes connected to the specified one by first and second level relationship. In addition, we allow user to choose maximum number of nodes that can be visualized and to filter the data based on a time range.

In order to facilitate navigation, user can click on any node to center the visualization around it or to go the corresponding profile page. Furthermore, names of people, documents and relationship types are displayed in visualizations. Additional metadata, such as dates or number of relationships are displayed on the side panel. We also keep track of hovering history, which allows user to easily locate previously visited nodes in the visualization.

Storytelling

Storyboard visualization gives user freedom to create visual narrative. Selected people and document nodes are linearly placed on the story board and user can arrange them according to his needs. He can view existing relationships, create new ones, remove items, enter title and description of the narrative, draw and modify shapes to emphasize important parts. Finally, such a visual narrative can be saved or exported as an image.

APPENDIX D
Client-Server Data Visualization in RoSE

(Online version of article in this appendix at: http://rosedocumentation.wordpress.com/2012/07/18/client-server-data-visualization-in-rose/)

This articles describes the overall architecture of the Data Visualizations of RoSE. The entire pipeline is covered here, from server initiation to client-side visualization. Pre-requisites for this article include a basic knowledge of Ruby on Rails, and the client-server model of web development.

Overview

The goal of the RoSE visualizations was to provide data a way to dynamically view data from a MySQL database edited and authors in a Ruby on Rails environment, and visualized in Flash. Thus, Ruby on Rails is the server-side framework, and Flash is the client-side visualization platform. The basic steps required for visualization are thus the same regardless of the data processing algorithms or specific visualizations, which will be covered later.

The communication starts in Ruby/Rails (actually, the web front end of Rails server, which is the online data entry and interface). Flash is invoked to start the visualization window, while the only data sent at this time is the id of the user or document to be visualized. Flash/Flare then creates a visualization, which generates a data request that goes back to the server. The server responds by building a data graph, and transmitting it in JSON format. Flash/Flare receives the data as a load event, decodes it, and builds a client-side representation that it uses to create sprites and layout nodes.

The full details of the process, and which specific functions in RoSE are called, is as follows:

SERVER
- Request made via ApplicationHelper::user_vis_link
- Popup window requested
- VisualController::index initiates visualization request
- Get host name
- Check requested server name
- Get default user & doc ids
- Invoke visual/index.html.erb view
- Start Flash visualization using swf_tag

FLASH
- Invokes changeVis (request visualization type)
- Initialize VisualizationParameters
- Set the user or doc id we wish to visualize.
- Overwrite defaults with Ruby/Rails flash vars for real server_name and user_id
- Set server URL from server name
- Create visualization object (timeline, network, document_circlePack, people, people_circlePack)
- Visualizations all have a loadData function (found in base class RoseVis)
- loadData does the following:
- Creates a URL loader
- Builds the API query (buildAPIQuery), and sends the request to the server
- Adds a listener to respond to the returned data stream

SERVER
- buildAPIQuery constructs the data query as:
- people: //{server}/data/user_network.json?kind=all&id={user_id}
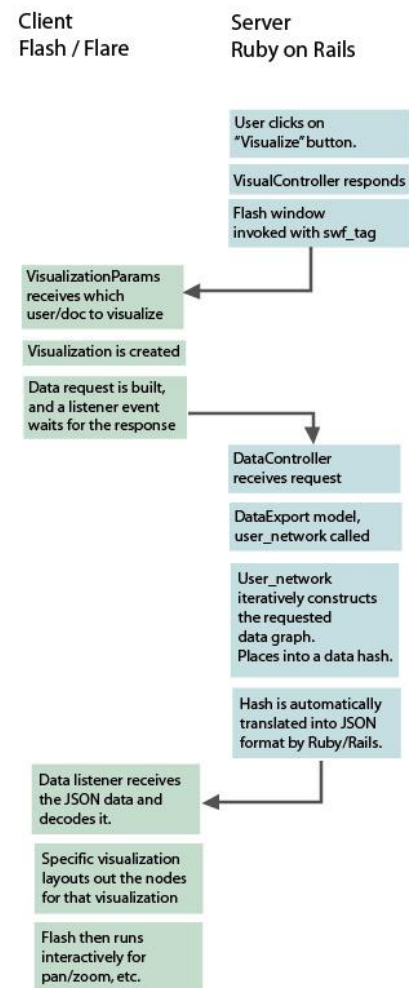
- documents: //{server}/data/doc_network.json?kind=all&id={doc_id}
- timeline: //{server}/data/simple_vis.json?kind=all&keyword={key_id}
- Rails DataController responds
- Maximum depth, doc and users are set from parameters
- Model function DataExport::user_network_data is called
- Iterative process builds the data graph
- Objects are translated in attribute-strings using specific toDataHash functions on each model.
- Graph is stored as a large hash of attribute-strings.
- Ruby/Rails does the translation from the hash into JSON text file.

FLASH
- loadData listener waits for data, then:
- decodes JSON data into raw_data (as Users, Docs, Rels)
- calls buildData (derived method), or displays data error — to parse into nodes
- calls visualize(data) to perform layout
- Flash then runs interactively with the nodes that were placed

These above steps are preformed regardless of the visualization, and also regardless of the algorithm used to build the data graph. A specific visualization provides additional methods by overriding the RoseVis:buildData function, and the RoseVis:visualize function, which determine how that unique visualization constructs data and lays out nodes (for example, network layout versus circle pack).

The data algorithm can also change. For example, the DataExport:user_network method generates a user-graph from a starting user, while the DataExport:doc_network generates a document-graph from a starting document. The DataExport:simple_vis_data, used by the timeline visualization, generates a set of nodes based on a keyword.

Original Method (2008)

To examine the specifics of how data visualizations are stored and created, I will rely on a simple example throughout this article. It is shown in the following figure:
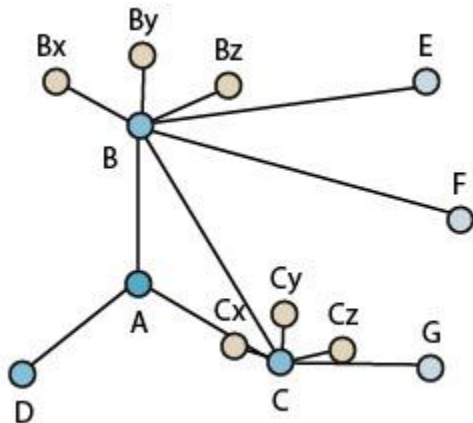
Figure 1. Sample user graph

The nodes A,B,C,D,E,F,G are people, and the nodes Bx,By,Bz and Cx,Cy,Cz are documents.
This is a person network visualization in Rose. However, the data principles here are applied to several visualizations (including circle pack), not just to this visualization.

The original method of data building, by Pehr Hovey, involved a recursive algorithm which starts at the root node A, and then did a breadth-first search of the user nodes. At each level, at the same time, the surrounding documents of each node are placed into the master node list.

Thus, the order of processing was:

A
  B (Bx, By, Bz)
  C (Cx, Cy, Cz)
  D
     E
     F
     G

Doing a breadth-first search allows the user to specific a max_depth to terminate the graph traversal. For example, a max_depth of 2 would stop after D (avoiding E,F,G). This is necessary in Rose because the graph connectivity can be arbitrarily large.

It is also necessary to maintain a seen_nodes list, to avoid repeatedly visiting nodes. For example, in the graph above then node C might be processed twice, first as a child of A, and then as a sibling of B. The fact that Rose graphs are not hierarchical means that nodes may be visited twice. However, this visualization method is hierarchical from a root, so we avoid repeated processing by recording which nodes were already visited.

This recursive algorithm was developed in 2008, when the main goal was to get RoSE up and running as early as possible. However, there are several inefficiencies in this strategy.

New Method (2012)

A new visualization strategy for Rose was developed in 2012 (R. Hoetzlein). The basic observation is that the recursive nature of the algorithm requires a large number of SQL requests to the server. First, the neighbors of each node are individually requested, then the documents around each node are individually requested. For example, the documents Bx,By,Bz around B result in mysql database request. However, this is done for each individual node. The number of requests made to the server is therefore: O(2N), where N is the average number of nodes to be visualized. Thus, if the visualization will contain 1000 nodes, this requires 1000 sql requests to the database. In addition, there is also overhead due to the seen_list.

Ideally, we would like the algorithm to make database requests independently of the number of nodes. The new strategy is an iterative one which considers a "wavefront" of active nodes. This wavefront represents the last visited nodes whose outgoing neighbors have not yet been found. The best way to understand this is via the following comparison:

Recursive breadth-first search   Iterative wavefront search
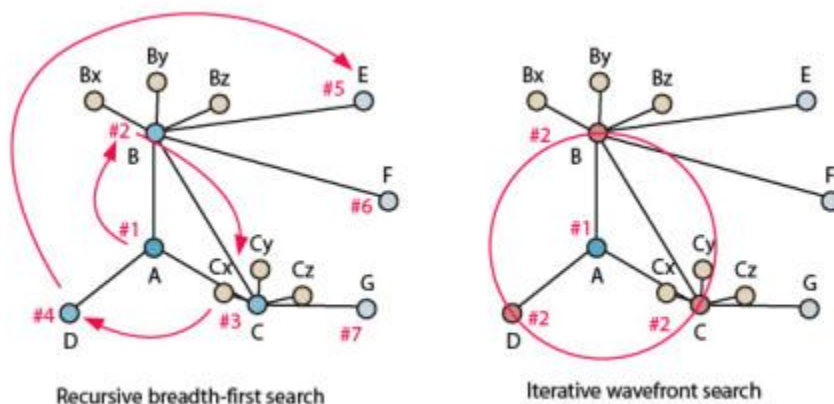
In this figure, the #s refer to the step in the process, which result in unique sql-database requests. In the iterative wavefront algorithm, the nodes B,C,D are each found using a single request. These nodes become the "active_front", and the node A is placed into the inactive final list.

In the next level of iteration, the active_front (B,C,D) is used as the seed for the next wavefront.
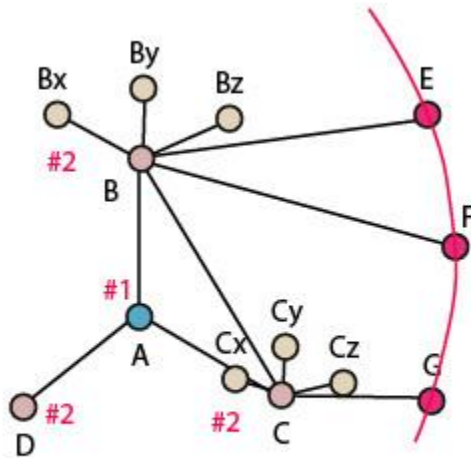
Iterative wavefront search

This generates the next level, E,F,G, using a single request to the database. The new active front is more clearly defined as: All those nodes which have a connection to any of the nodes in the previous active front, and which are not in the inactive list. So, E,F,G are found as neighbors to B,C,D (A is ignored because it is now inactive). In effect, the algorithm can be visualized as a wave which travels outward from the center node.

In Ruby/Rails, this is done in only 5 lines of code as follows:

```
for level in 0..max_level

    # find all relationships outgoing from active people
    new_edges = Relationship.find(:all, :conditions =>['src_user_id IN (?)', @active_nodes.map{|p| p.id} ] )
    @final_edges << new_edges

    # find all connected people associated with relationships
    new_nodes = User.find( new_edges.map{|r| r.dest_user_id} )
    @final_nodes << new_nodes

    @active_nodes = new_nodes
end
```

The key here is the line which says "new_edges = Relationship.find (:all, :conditions=>['src_used_id IN (?)', @active_nodes.map{|p| p.id}).
This asks the database for all relationships (edges) in which one side of the edge matches any node in the previously active wavefront. It retrieves the neighbors of B,C and D in a single step. The @active_nodes.map{|p| p.id} is done because the SQL query needs the ids, while the active_nodes stores the actual Rails object. This map command just translates from a list of objects into a list of ids. What is returned is a list of Relationships (joining Users to Users).

The next line finds the nodes at the end of these edges:
"new_nodes = User.find(new_edges.map{|r| r.dest_user_id}) "
This is a basic find request, which asks for all the Users (nodes) which are given by a list of ids. The new_edges is a list of Relationship objects, so the map again is used to convert to a list of requested ids. Note that the ids requested are the "dest_user_id", or the other end of the edge.

That's it. The found nodes and edges are recorded in the final_list, and the active_nodes are then set to the new nodes just found.

The run-time of this method is O(2L), where L is the number of levels, not the number of nodes. In a typical graph of 1000 nodes, the max_depth needed to reach them might be 3. Therefore, the number of mysql database requests is 6 instead of 6000.

In computer science terms this iterative wavefront method, which finds a hierarchy of nodes embedded in a graph, is basically a parallel breadth-first search (BFS) which has been optimized for SQL databases. Expressed as a tree (non-cyclical), all the nodes at a given level of the tree are retrieved in a single database operation.

Data Transmission

The above discussion refers to the data processing, or Ruby/Rails side of the visualizations.
From both the original and new algorithm, the result is a set of nodes and edges to be visualized. This data must then be packaged for transmission to Flash client.
Originally, the packaging was done for indvidual users, docs, and relationships, in the DataExport:user_network_data method follows:

@data[:users]= @node_user_list.compact
@data[:docs] = @node_doc_list.compact
@data[:rels] = @node_r_list.compact

This produces a hash, in which the users, docs and relationships are separated.
The resulting JSON data transmitted to the server looks like this:

{"es":
[{"type":"Y","src":"\d3746","start":null,"dst":"\d1764","t_type":"Influence","end":null,"id":"\5463"},
{"type":"Y","src":"\d3746","start":null,"dst":"\d1807","t_type":"Influence","end":null,"id":"\5406"},
{"type":"Y","src":"\d3746","start":null,"dst":"\d1800","t_type":"Influence","end":null,"id":"\5446"},
{"type":"Y","src":"\d3746","start":null,"dst":"\d1720","t_type":"Influence","end":null,"id":"\5432"},
{"type":"Y","src":"\d3746","start":null,"dst":"\d1707","t_type":"Influence","end":null,"id":"\5601"},
{"type":"Y","src":"\d3746","start":null,"dst":"\d1764","t_type":"Influence","end":null,"id":"\5508"},
{"type":"Y","src":"\d3746","start":null,"dst":"\d1764","t_type":"Influence","end":null,"id":"\5509"},
{"src":"\d3746","x_type":"Udr","dst":"doc265578","t_type":"Created","id":"udr14787"},
{"src":"\d3746","x_type":"Udr","dst":"doc265580","t_type":"Created","id":"udr14891"},
{"src":"\d3746","x_type":"Udr","dst":"doc265587","t_type":"Created","id":"udr14917"},
{"src":"\d3746","x_type":"Udr","dst":"doc265594","t_type":"Created","id":"udr14921"},

"users":[{"first":"Rama","levels":[3],"b_date":"1975-12-13","last":"Hextaban","kind":"historic","d_date":"2010-01-01","b_year":1975,"id_dx":
[1667],"\d671","udr1677","udr1602","udr1676","udr1167E","udr1118","udr11884","W6614","W6678"],"doc_count":8,"d_year":2010,"id":"\i1","id_count":3,"h_type":"user"},
{"first":"Javier","levels":[2],"b_date":null,"last":"Matias","kind":"","d_date":null,"b_year":"","id_dx":
[1614","W664","W1547","W2346","W2064","W3055","W4526","W4564","W5551"],"doc_count":3,"d_year":"","id":"udr1706","id_count":3,"h_type":"user"},
{"first":"John","levels":[2],"b_date":null,"last":"Stadler","kind":"","d_date":null,"b_year":"","id_dx":
[149","W7","W5","udr318","udr180","udr227","udr187","udr208","udr196","udr181","udr241","W12","W565","W678","W177","W2082","W3606","W3728","W4114","W5056","W6521],
{"first":"Steven","levels":[2],"b_date":null,"last":"Wilhauser","kind":"","d_date":null,"b_year":"","id_dx":
[udr5528","udr5320","2620","W3028","W1239","W1888","W2318","W2809","W3884","W4288","W4784","W5628"],"doc_count":2,"d_year":"","id":"udr1728","id_count":3,"h_type":"user"},
{"first":"John","levels":[2],"b_date":null,"last":"Dorner","kind":"","d_date":null,"b_year":"","id_dx":
[US7","U29","U30","U31","U17","W1887","W5524"],"doc_count":0,"d_year":"","id":"udr3765","id_count":8,"h_type":"user"},
{"first":"T. B.","levels":[2],"b_date":null,"last":"Eliot","kind":"","d_date":null,"b_year":"","id_dx":
[W1178","V1184","V1188","V1189","V1183","V1153","V1184","V1199","V1177","V152","V1157","udr1821","udr1931","W54","W308","W1573","W1887","W3187","W3830","W3738","W4751","W4508","W5521],
{"first":"Lindsay","levels":[1],"b_date":null,"last":"Thomas","kind":null,"d_date":null,"b_year":"","id_dx":
[W5864","W5005","W5000","W5007","W6671],"doc_count":0,"d_year":"","id":"\i1405","id_count":8,"h_type":"user"}],

"docs":[{"num_edges":1,"levels":[2],"vol":0,"date":"1963-11-12","name":"Rabbit, Run","h_type":"Novel","id_dx":
[Xudr218"],"doc_count":0,"dp":1963,"ref_user_count":1,"id":"doc254406","num":0,"h_type":"doc"},
{"num_edges":1,"levels":[2],"vol":3,"date":"2006-01-01","name":"Terrorist (novel)","h_type":"Novel","id_dx":
[Xudr485"],"doc_count":0,"dp":2006,"ref_user_count":1,"id":"doc254407","num":0,"h_type":"doc"},
{"num_edges":1,"levels":[2],"vol":1,"date":"","name":"Too Far to Go","h_type":"Book","id_dx":
[Xudr221"],"doc_count":0,"dp":"","ref_user_count":1,"id":"doc254413","num":0,"h_type":"doc"},
{"num_edges":1,"levels":[2],"vol":1,"date":"1986-01-01","name":"Rabbit at Rest","h_type":"Novel","id_dx":
[Xudr187"],"doc_count":0,"dp":1986,"ref_user_count":1,"id":"doc254428","num":0,"h_type":"doc"},
{"num_edges":1,"levels":[2],"vol":0,"date":"1981-09-02","name":"Rabbit Is Rich","h_type":"Novel","id_dx":
[Xudr208"],"doc_count":0,"dp":1981,"ref_user_count":1,"id":"doc254429","num":0,"h_type":"doc"}],

"results":
{"total_nodes":58,"total_docs":45,"actual_depth":2,"total_users":13,"max_depth":2,"total_rels":55},

"request":
{"format":"json","kind":"all","token":"12345","action":"user_network","id":"1","controller":"data"}}

There are several things to notice about this data. First, users and documents are separated. This seems natural, but actually both can be considered nodes in a graph. The new JSON format, below, has only two groups, nodes and edges. Secondly, each node includes a list of its surrounding relationships ("udr" and "Xudr"), as can be seen in the JSON listing above. This is necessary, since the radial visualization must know what nodes are around it in order to properly layout the surrounding edges. However, it is not necessary to transmit from server-to-client because this uses a great deal more bandwidth than necessary. All the information needed to recreate the neighboring nodes can be found in the relationships list alone. This adds more code to the client-side, but greatly improves download performance.

Here is the new JSON format for visualizations:

{"edges":
        [{"type": "r", "src": "u1", "rt": "Friend Of", "dst": "u83746", "id": "r6679"},
        {"type": "r", "src": "u1", "rt": "Colleague Of", "dst": "u91406", "id": "r6671"},
        {"type": "r", "src": "u83746", "rt": "Influences", "dst": "u81704", "id": "r5509"},
        {"type": "r", "src": "u83746", "rt": "Influences", "dst": "u81715", "id": "r5523"},
        {"type": "r", "src": "u83746", "rt": "Influences", "dst": "u81725", "id": "r5492"},
        {"type": "r", "src": "u83746", "rt": "Influences", "dst": "u81753", "id": "r5524"},
        {"type": "udr", "src": "u81715", "rt": "author", "dst": "doc254417", "id": "udr221"},
        {"type": "udr", "src": "u81715", "rt": "author", "dst": "doc254428", "id": "udr187"},
        {"type": "udr", "src": "u81715", "rt": "author", "dst": "doc254429", "id": "udr208"},
        {"type": "udr", "src": "u81715", "rt": "author", "dst": "doc254430", "id": "udr195"},

"nodes": [{"type": "u", "dd": "2010-01-01", "name": "Rama Hoetzlein", "id": "u1", "d": "1975-12-15"},
        {"type": "u", "dd": "1616-04-23", "name": "William Shakespeare", "id": "u83746", "d": "1564-04-26"},
        {"type": "u", "dd": null, "name": "Lindsay Thomas", "id": "u91406", "d": null},
        {"type": "u", "dd": null, "name": "Javier Marias", "id": "u81704", "d": null},
        {"type": "u", "dd": null, "name": "John Updike", "id": "u81715", "d": null},
        {"type": "u", "dd": null, "name": "Steven Millhauser", "id": "u81725", "d": null},
        {"type": "u", "dd": null, "name": "John Donne", "id": "u81753", "d": null},
        {"type": "d", "rtype": "Novel", "name": "The Witches of Eastwick", "id": "doc254435", "d": "1984-04-12"},
        {"type": "d", "rtype": "Short Story", "name": "Trust Me (book)", "id": "doc254436", "d": ""},
        {"type": "d", "rtype": "Novel", "name": "Rabbit Redux", "id": "doc254441", "d": "1971-01-01"},
        {"type": "d", "rtype": "Film", "name": "The Illusionist", "id": "doc259548", "d": ""},
        {"type": "d", "rtype": "Collection", "name": "The Barnum Museum", "id": "doc259555", "d": "1990-06-01"},
        {"type": "d", "rtype": "Play", "name": "The Cocktail Party", "id": "doc256133", "d": ""},
}

The amount of data is greatly reduced, due to the following changes:
1. Only two major blocks: nodes and edges
2. Surrounding relationships are not included with nodes. Instead, these are now derived from the edges in the Flash code.
3. Nodes are distinguished by abbreviated types, u=user, d=doc. This saves 3 to 4 bytes per node.
4. Other abbreviations. d=document or birth date, dd=death date. Saves several bytes per node.

The new Ruby/Rails code for packaging JSON data is as follows:

```
# Convert data to JSON hash format
@data[:nodes] = @final_nodes.flatten.uniq.map{|r| r.toDataHash}
@data[:edges] = @final_edges.flatten.uniq.map{|r| r.toDataHash}
```

The final list of nodes and edge objects are flattened to remove sub-arrays, and uniq is called to make sure there are no duplicates. Finally, the entire list of objects is converted attribute-strings using toDataHash.
This is a virtual function on each RoSE model which takes in an object, and produces a hash. For example, here is the hash for users:

```
def User.toDataHash(u)
{:id=>u.visID, :name=>u.full_name, :type=>"u", :d=>u.birth_date, :dd=>u.death_date }
end
```

A user object 'u' is converted into a set of names attributes. Similar toDataHash functions exist for all the model objects that can be visualized. Ruby/Rails will automatically take these hashes and generates the actual JSON text file.

Flex/Flare Visualization

The final step is to visualize the resulting data. In July 2012, this part has undergone the most amount of revision. Prior to 2012 the RoSE visualizations consisted of a number of separate modules that each had their own loadData method, server response, data building, and layout. However, it was realized that this results in very repetitive coding, since many of these steps are identical in all visualizations. For example, the network, people, people_circlepack, and document_circlepack visualizations all have the same loadData function (see overview above). Their buildData functions all start by constructing a master list of allnodes and alledges.

In 2012, this functionality was incorporated into a single base class RoseVis. All other visualizations, including the timeline and storyboard, are now derived from the RoseVis class. The RoseVis provides a central location for storing VisualizationParameters, the master list of all raw nodes and edges, color preferences, and references to layouts. It now provides a number of helper API functions for generating graph data regardless of layout.
These functions include:

- BuildAllData – Constructs a master list of all nodes & edges
- BuildLevels – Constructs a subset, which is a hierarchical layout based on a central node
- BuildRings – Constructs the surrounding nodes, such as documents around a person
- BuildEdges – Constructs the connecting edges given a set of nodes
- SetEdgeColors – Sets the edges to their default RoSE colors based on type
- MakeLabels – Attaches text labels to each node based on name
- FindNeighbors – More low-level, finds the neighbors of a node. Used for highlighting, etc., and replaces the neighbor relationships found in the older JSON format.
- FindEdges – More low-level, finds the edges of a given node by a particular edge type.
- FindNode – More low-level, finds a particular Flex/Flare node object given its JSON id.
- setRadial – Indicates the type of objects that are primary nodes (placed in levels)
- setSurround – Indicates the type of objects that are secondary, around the primary

This strategy dramatically changes the way visualizations are created. For example, the older buildData method for radial visualizations consisted of hundreds of lines of code.

The new radial visualization buildData does the following:

```
protected override function buildData() : Data {

    active_data = new Data(true);

    // Set radial and surround types for hierarchy-type visualization
    setRadial ( "u", visParams.max_users );
    setSurround ( "d", visParams.max_docs );

    // Build level hierarchy of nodes & edges
    // - user_id is root of hierarchy (center of graph)
    // - "r" defines the outgoing edges
    // This sets the "level" of each node at consecutive circles (used for layouts)
    BuildLevels ( active_data, "u"+visParams.user_id, "r" );

    // Build rings around each node.
    // - Nodes are taken from incoming data
    // - "udr" defines the surround edge type
    // This sets the "cnt_child" and "parent" of each node (used for layouts)
    BuildRings ( active_data, "udr" ) ;

    // Create edges and color them
    BuildEdges ( active_data, "r");
    BuildEdges ( active_data, "udr");
    SetEdgeColors ( active_data );

    // Attach default labels to all nodes
    MakeLabels ( active_data );

    return active_data;

}
```

The first line create a new subset of data called the active_data. This is not necessarily all the data which was loaded via JSON, it is a subset for the current # people and # documents requested.

The next two lines indicate this is a radial visualization with users "u" as primary, and documents "d" as secondary (surrounding). Note that the radial/surround are not specific to the network visualizations. These functions would also be used in circle packing to distinguish the primary nodes (large circles), and secondary (inside circles). The difference is only in layout, while setRadial and setSurround should be considered as data-processing requests.

The next line, BuildLevels, tells the RoseVis class to take all data and construct a subset called active_data, who root node is given by the user_id visParam, and whose edges are based on the "r" (relationship edge type). It does not specify which nodes surround these user nodes.

The next line, BuildRings tells the RoseVis to also add nodes based on the "udr" (user-document relation), which places documents around all the users. At the same time, it updates the cnt_child and parent, which are needed to properly position nodes during layout.

The next three lines, BuildEdges and SetEdgeColors, connect the nodes with edges. These request edge-lines for all user-to-user relationships ("r"), and for all user-to-doc relationships ("udr").

Finally, the last line MakeLabels asks RoseVis to attach text labels to all of the nodes in the graph.

It is important to note that nearly all of these steps are optional. The only required step in the above setup is to setRadial/setSurround and BuildLevels. Adding rings of documents, adding edge-lines, and making labels are all optional. These are left out when doing a circle packed visualization, for example.
In other visualizations, such as timelines or storyboards, even the required BuildLevels would be replaced by a different function such as BuildKeywords (not yet implemented).

Dynamic Layout

A key feature of this new strategy is the ability to construct sub-sets of a full dataset and visualize those without having to contact the server again. A JSON request returns a complete graph, which may have thousands of nodes. The BuildAllData function places this into global allnodes, alledges variables. This data is only change when a new user, document or visualization is selected.

Using the Filter panel in the RoSE visualizations, it is possible to select a different number of # people or # documents. For interactivity, this does not cause a new JSON request to be performed, but instead calls the RoseVis.refresh_layout function. This does the following:

```
public function refresh_layout():void {
    // NOTE: This function assumes BuildAllData already loaded all JSON data.
    // The following code processes a subset of all data to produce a new layout interactively

    // Clear children (text objects) from global node data
    ClearChildren ();

    // Flare clear data
    active_data.clear();

    // Visualization-specific handling of data (derived method)
    active_data = buildData ();

    vis.data = active_data;

    // Visualization-specific layout update (derived method)
    updateLayout();
}
```

ClearChildren removes any text objects which were previously attached to node sprites.
The active_data.clear function invokes Flare to empty the nodes from the current dataset. The full set of nodes still exists in the allnodes/alledges lists.
The buildData function is now called again for a particular visualization, which first allocates a new active_nodes array. The active_nodes is a changing subset of the full data. When buildData calls BuildLevels, for example, it populates the active_nodes from the allnodes data without requiring new JSON data from the server. This allows for a dynamic subset of the data to be created interactively.
The data is replaced in the visualization, without unloading the visualization itself, by doing "vis.data = active_data"
Finally, updateLayout is called on the specific visualization to reposition the nodes. The sole purpose of updateLayout is to change the x/y, or size of a set of nodes.

The following figure shows the dynamic layout feature being used to change the # of documents in the Flash interface.
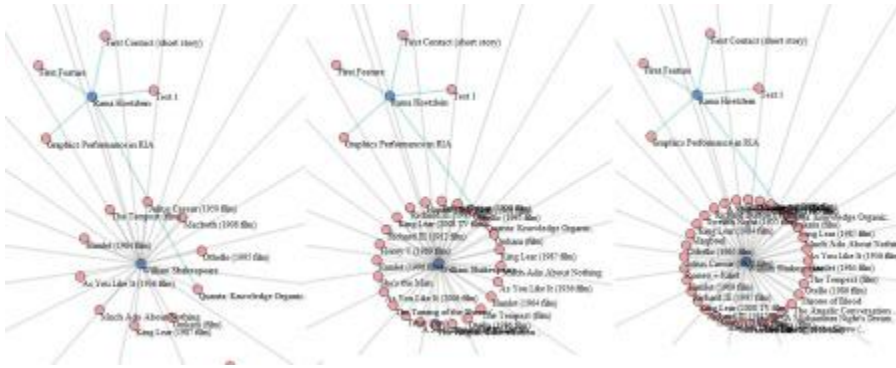
Additional Changes

The RoseVis base class provides some other new functionality which was previous found in each visualization.
These are as follows:

1. updateLayout original contained both pan/zoom and node placement code. Now, the sole responsibility of updateLayout is node placement (and color/size). Pan/Zoom can happen without requiring nodes to be placed again.
2. updateZoom is a new function in a layout which does pan/zoom. The method of pan-zoom has changed so that it now modifies the transformation matrix of the parent vis sprite. This greatly improves performance as nodes can now be laid out in a static world coordinate space, and simplifies code as pan/zoom are no longer a part of node layout. Some node adjustments may still be desirable in updateZoom, such as showing or hiding text labels based on the zoom level.
3. Layouts, such as RadialNodeLayout, now have access to global variables via the RoseVis base class. For example, radialMax, surroundMax, and spacing, which define the angular spacing in the radial network visualization are retrieved by saying: roseVis.radialMax and roseVis.surroundMax.
4. Data which is a part of VisualizationParameters can also be accessed in each visualization. For example: roseVis.visParams.spacing, roseVis.visParams.width, roseVis.visParams.height

APPENDIX E
Observer's Report on Classroom Demo of RoSE
[Student names have been blacked out due to RoSE human subjects protocol]

Monday, May 21st, 2012

The atmosphere was casual as some 20 students began milling into the Transcriptions Lab, logging onto computer terminals, opening laptops, and entering the brave new world of UCSB's Research-oriented Social Environment. Humanities developers Dana Solomon and Lindsay Thomas introduced the tasks at hand: to register a user profile, compile a collection of people and documents related to a research project, to add new entries to the RoSE database, and to work towards a visual storyboard that would present relationships between people and texts in a new light.

The process, at first, was slow going. Confronted with an unfamiliar interface, students gazed perplexed at a number of options and buttons, clicking here and there to see what might happen. Some found themselves opening new tabs, playing a game of bubble spinner, or looking at the latest fashions on chloë.com, all par for a course that is entitled, "Distracted Reading."

Gradually though, progress was made. As students began figuring out ways to add items and create collections, ejaculations of "cool" and "aha" began to circle. When a student figured out a function, he/she was quick to explain the move to surrounding peers. This collaboration between students to learn the system and work around areas of low functionality was especially encouraging, and generated a sense of camaraderie to keep exploring the RoSE site.

A few minutes into the exercise, Joshua Dickinson entered the room, and upon his introduction as a lead programmer, murmurs of interest and appreciation rippled through the crowd.

An initial stumbling block was an error in the Collections functionality. The team reached out to lead programmer Rama Hoetzlein, who in fact corrected the error in real time, allowing the trial to gain momentum and proceed smoothly.

Students were especially excited by the prospect of adding new keywords and items. One student, ▮▮▮▮▮▮▮, began adding a number of vintage advertisements from the 50s and 60s to the RoSE database. He appreciated the ability to add images of the advertisements to the document pages. When asked why he chose advertisements, he explained that "Advertisement" was alphabetically the first type of item that one was allowed to add.

Another student, ▮▮▮▮▮▮, explored the ability to search and find different people and documents. She expressed some confusion over the different search box options, but eventually ended up searching through the "People" and "Documents" pages respectively. She was impressed by the ability to search and find different people and documents, explaining, "It feels like a giant bibliography."

Students had a lot of fun playing around with the "Relationships" feature. One studentmischievously married Professor/RoSE supervisor Rita Raley to the novelist Henry James.Other students called to friends across the room, "I'm your rival" and "I'm a scholar of yourwork."

As students began to have success creating their collections, some began exploringthe "Storyboard" function. The different shapes and colors were a big success, with students making intricate webs that--if unclear in academic significance--proved aesthetically interesting. ▮▮▮▮▮▮ noted that it was working well for making venn diagrams.

As the class period drew to a close. Students reflected on their experience of RoSE. Despite some difficulties with the interface, many students saw potential for the nascent site.

▮▮▮▮▮▮ noted that the facts and information in the database were less interesting than the way users could explore how things fit together. For the site to be viable, he thought a number of users would have to be actively importing user-generated content, noting that it would be "up to them" to decide what connections and information the system would reveal.

▮▮▮▮▮▮ thought that RoSE might be especially successful with "a younger generation who have grown up in the digital age and are even more connected to the internet." She saw potential to use RoSE in high school classes.

At the end of the hour, students logged off and shifted their gazes to the outside world. Student ███████ seemed to capture the prevailing mood, as he said, "Now that I know how to play around, I just need to figure out how I want to use it."

Student Consent Form for RoSE Use-Scenario Study

RoSE: Bibliographical Knowledge as Social Knowledge
Consent Form
Approved for use through: 4/4/2013

PURPOSE:
You are being asked to participate in a research study. The purpose of the study is to determine the usability of the RoSE (Research-oriented Social Environment). RoSE is a system for exploring the humanities that encourages you to seek out relationships between authors, works, and commentators--living and dead--as part of a social network of knowledge. The study will determine what students and researchers in the humanities want from a humanities research social environment, as well as which features of RoSE are easy to use and which aren't.

PROCEDURES:
If you decide to participate, you will fill out a questionnaire after completing specific tasks in RoSE. The purpose of the questionnaire is to determine the viability of RoSE as a new tool for humanities research and classroom exploration and the general usability of RoSE as a system. You will only be asked to fill out one questionnaire.

The time commitment for filling out this questionnaire is approximately 15-20 minutes. The study itself will be performed over approximately two weeks in May of 2012 and will include a classroom use study and a conference use study. Approximately 50 subjects will be involved.

RISKS:
We do not anticipate any mental or physical health risks as a result of participating in this study.

BENEFITS:
While we expect the development of RoSE to improve the ways in which scholarly resources in the humanities are collected, stored, and shared, there is no direct benefit to you anticipated from your participation in this study.

CONFIDENTIALITY:
The data we collect will be used in writing the final report and white paper for the National Endowment for the Humanities (NEH), the agency funding the development of this project. The white paper will be published online on the NEH public web site). We will report the results of the usability questionnaires; the purpose of this disclosure will be to detail what improvements to RoSE we have made over the funding year and to report user feedback on RoSE. If you create a profile in RoSE, the name you use for that profile - which can be any name of your choosing - will be visible in the system, but we will not use your name or any personal information in our reports.

RIGHT TO REFUSE OR WITHDRAW:
You may refuse to participate and still receive any benefits you would receive if you filled out the questionnaire. You may change your mind about filling out the questionnaire and quit after you have started.

QUESTIONS:
If you have any questions about this research project or if you think you may have been injured as a result of your participation, please contact:

Professor Alan Liu, ayliu@english.ucsb.edu

If you have any questions regarding your rights and participation as a research subject, please contact the Human Subjects Committee at (805) 893-3807 or hsc@research.ucsb.edu. Or write to the University of California, Human Subjects Committee, Office of Research, Santa Barbara, CA 93106-2050

PARTICIPATION IN RESEARCH IS VOLUNTARY. YOUR SIGNATURE BELOW WILL INDICATE THAT YOU HAVE DECIDED TO PARTICIPATE AS A RESEARCH SUBJECT IN THE STUDY DESCRIBED ABOVE. YOU WILL BE GIVEN A SIGNED AND DATED COPY OF THIS FORM TO KEEP.

Signature of Participant or Legal
Representative:＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿Date:＿＿＿＿＿＿＿Time:＿＿＿＿

<u>**APPENDIX G**</u>
RoSE User Questionnaire

This questionnaire gives you an opportunity to tell the RoSE development team at UCSB about your reactions to RoSE. After asking you for some initial information, the questionnaire is divided into *conceptual* and *functional* questions about the system. Your responses will help us understand what aspects of the system we should focus on in further development.

NOTE: This system is currently in beta (still in development) and is not yet fully polished. We are aware that not all features are fully functional. Please take this into consideration when giving us your answers.

--Thanks from the RoSE team (faculty and students at UCSB working on a NEH Digital Humanities Start-up grant)

Your Information

The information you provide in this section will help us learn more about potential users of RoSE. All of your information will remain anonymous.

1. Gender:

    ----

2. Major:

    ----

3. How many hours per week do you estimate you use social networking sites (e.g., Facebook, Twitter, Flickr, Goodreads, Pinterest, etc.)?

    ___ 0-5 hrs/week
    ___ 5-10 hrs/week
    ___ 10-15 hrs/week
    ___ 15-20 hrs/week
    ___ 20 + hrs/week

    ----

4. How familiar are you with the following online bibliographic tools? Please rank each site on a scale from 1-7 according to your familiarity.

VERY
UNFAMILIAR   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **VERY**
FAMILIAR

    ___ Google Scholar
    ___ Library of Congress
    ___ Melvyl
    ___ Pegasus

___ LibraryThing
___ Zotero
___ Worldcat

Conceptual Questions

RoSE is a system that encourages you to perceive, navigate, and interact with the bibliography of humanist knowledge as if it were a social network in which everyone and everything—past and present—were interconnected in a community of knowledge. RoSE seeds its system with a set of authors and documents harvested from online databases but then allows you to evolve the system by adding new authors, documents, and relations.

Please read each statement below and answer the question as directed:

5. By fusing a "library" and "social network" model of knowledge, RoSE adds *a valuable way* to explore ideas beyond what a normal library or social network system would allow. Indicate how strongly you agree or disagree by circling the corresponding number on the scale.

STRONGLY   **STRONGLY**
AGREE   1   2   3   4   5   6   7   **DISAGREE**

COMMENTS:

----

6. How do you think RoSE compares to the following sites in terms of doing research? Please rank each site on a scale from 1-7. *If you have not used a particular site, please leave that space blank.*

ROSE IS MUCH   **ROSE IS MUCH**
WORSE   1   2   3   4   5   6   7   **BETTER**

___ Google Scholar
___ Facebook
___ Worldcat
___ Project Gutenberg
___ Library of Congress
___ Melvyl
___ Pegasus

----

7. The "history track" feature for keeping a record of your movement through RoSE is valuable.

STRONGLY   **STRONGLY**
AGREE   1   2   3   4   5   6   7   **DISAGREE**

COMMENTS:

----

8.   The "collection" feature for keeping a collection of materials gathered from RoSE is valuable.

STRONGLY   **STRONGLY**
AGREE   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **DISAGREE**

COMMENTS:

----

9.   The "storyboard" feature for keeping a collection of materials gathered from RoSE is valuable.

STRONGLY   **STRONGLY**
AGREE   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **DISAGREE**

COMMENTS:

----

10. What do you think you would most likely use RoSE for when the system is finished?

----

11. Who do you feel would be most likely to use RoSE? Please check all that apply.

      ___ Students
      ___ Instructors (for classroom use)
      ___ Individual researchers
      ___ Other

      IF OTHER, PLEASE EXPLAIN WHO:

----

12. Being able to take an active role in evolving RoSE by adding authors and documents as well as relations between them helps you learn more about a movement or idea than just passively browsing existing content.

STRONGLY   **STRONGLY**
AGREE   1   2   3   4   5   6   7   **DISAGREE**

COMMENTS:

----

13. Rank the following features of RoSE on a scale from 1 to 5 according to what you found most useful for performing your tasks in RoSE.

1 = Most important; 2 = Important; 3 = Neutral; 4 = Unimportant; 5 = Very unimportant; N/A = Not Applicable

   ___ Profile pages for authors and documents
   ___ Data visualizations
   ___ History tracks
   ___ Collections
   ___ Storyboards

----

14. Which of the following do you think is the **MOST IMPORTANT** feature of RoSE?

   a. The ability to use RoSE to conceptualize networks of relationships between users, authors, and documents
   b. The ability to use RoSE to add authors, documents, and relationships to preexisting bibliographical data
   c. The ability to use RoSE to visualize social networks
   d. The ability to use RoSE to create collections of bibliographical information
   e. The ability to use RoSE to participate in shared collections
   f. The ability to use RoSE to build an understanding of an author, movement, or other phenomenon that can be shaped into an argument or story (e.g., through the storyboarding feature)

----

15. Which of the following do you think the developers of RoSE should concentrate on as the single most important way to make RoSE a better way to learn, explore, and think about knowledge?

   a. Improving and adding features to the "library" aspect of RoSE (for example, by providing fuller citations for works, direct links to the full texts of works or excerpts where available, etc.)
   b. Improving and adding features to the "social network" aspect of RoSE (for example, by allowing for Facebook-like interactions and chat between users, by automating the profile pages of historical authors and works so that they seem to react dynamically to changing conditions or today's news [e.g., quotations from his works appearing on Shakespeare's page in reaction to current events], etc.)
   c. Improving and adding features to RoSE's visualizations

----

16. What do you think overall of the RoSE concept (separate from any current technical or functional issues)?

Functional Questions

Please read each statement and indicate how strongly you agree or disagree with the statement by circling the corresponding number on the scale. Please write comments to elaborate on your answers if needed.

17. Overall, I am satisfied with how easy it is to use RoSE.

STRONGLY   **STRONGLY**
AGREE   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **DISAGREE**

COMMENTS:

----

18. The help information that comes with the system ("What You Can Do With RoSE" help section) is effective in assisting me in completing the assigned tasks in RoSE.

STRONGLY   **STRONGLY**
AGREE   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **DISAGREE**

COMMENTS:

----

19. "Collections" are easy to use and manage (subscribe/unsubscribe to, add/delete new documents, etc.).

STRONGLY   **STRONGLY**
AGREE   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **DISAGREE**

COMMENTS:

----

20. The "storyboard" interface is easy to use.

STRONGLY   **STRONGLY**
AGREE   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **DISAGREE**

COMMENTS:

----

21. I thought it was easy to move from one function of RoSE to another (i.e., to move from adding people and
     documents, to tracking a history, to viewing collections, to visualizing a person's connections, to storyboarding, etc.)

STRONGLY   **STRONGLY**
AGREE   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **DISAGREE**

COMMENTS:

----

22. I felt confident in my ability to use the "history tracking" feature.

STRONGLY   **STRONGLY**
AGREE   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **DISAGREE**

COMMENTS:

----

23. The visualizations in RoSE are easy to use and understand.

STRONGLY   **STRONGLY**
AGREE   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **DISAGREE**

COMMENTS:

----

24. I used the visualization feature to display known relationships more than I used it to discover new relationships.

STRONGLY   **STRONGLY**
AGREE   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **DISAGREE**

COMMENTS:

----

25. "Keywords" are a useful way to organize and view the data in RoSE.

STRONGLY   **STRONGLY**
AGREE   **1**   **2**   **3**   **4**   **5**   **6**   **7**   **DISAGREE**

COMMENTS:

----

26. Overall, what functional feature of RoSE do you think most needs improvement, and why?

APPENDIX H
Student Suggestions for RoSE Developers
Excerpted from student questionnaires and online reflections for their course assignment in RoSE use-scenario study

From Student Reports: *[Student names have been blacked out due to RoSE human subjects protocol]*

███████:

"Ideally, I'd like to see the storyboards becoming more telescopic and interactive. I would love to be able to click on keywords like "Satire" and see a list of related quotes and/or authors appear."

"Authors like Danielewski and Bechdel constantly make reference to other iconic pieces. How cool would it be if we could have a direct link to those texts (or at least excerpts) while just looking at a text through RoSE. Of course this can't really work with current copyright statutes, but I just like to fantasize about the possibility of parsing a text, essentially doing a machine-assisted reading and jumping immediately to other machine-assisted readings of related texts."

"Also it would be really neat if there were a grid/snap-to feature on the storyboard"

███████:

"I would like to see the use of images in RoSE greatly increased. In this time of picture blogs and Facebook, images are a crucial commodity and attraction. If RoSE could capitalize on this, it would be a far more enjoyable experience."

███████:

"As far as improvements, I think the biggest addition would be short biographies for authors and summaries for documents similar to the way Wikipedia is set up. I think this would really add to the research portion of RoSE by giving it a little more depth. As it currently stands, it is a little difficult to differentiate between authors or documents with similar names."

"I think the storyboard could be improved with some small changes such as more color options, being able to adjust the size of the canvas and of the nodes, being able to edit and abbreviate the text that appears with the nodes, and of course being able to save."

"As for the collection feature, I think it would be helpful if you could see which collections a particular document is part of when looking at its individual page."

"I think being able to message other users is a good idea because it enables group work on the site and would make it easier for researchers to coordinate their efforts."

"One of the biggest possible pitfalls for RoSE is something that I believe has already been discussed, which is the potential for abuse. This seems like something that can easily be fixed, either through limiting some of the editing power to a few administrators, or maybe through some sort of system where accounts are tracked and any user found making repeated abuses of the system has their ability to edit revoked."

███████:

"The first problem I had was attempting to add articles which were not on the database yet. I would use the straight-forward "Add Document" tab only to discover after repeatedly typing in the necessary information, that the document failed to be added. After some time, I figured out how to add the authors of the articles, write their relationship to a document, and then add the document. I think this process is extremely tiresome and occupied too much of my time."

"The last problem I had was regarding actually making the storyboard. When my collection finally contained all the articles I desired, I clicked the "storyboard" option. I was then directed to a page where RoSE asked me to "select from my collections" in order to drag the documents onto the storyboard plane. However, there were no collections in that space. If it wasn't for spending around an hour repeating the process of trial-and-error, I would have never discovered that one must go onto one's own user page and make a storyboard that way. I think the website would be greatly improved if the option to make a storyboard actually worked."

███████:

"The first issue that I came up against in RoSE was just trying to understand and comprehend what it is exactly that I was supposed to do with it. If the emphasis of RoSE was supposed to be on the relationships between texts, authors and ideas how was it possible to illustrate these without giving a description? I personally felt that too much text on the page would be distracting and too little text on the page would leave to much room for interpretation of what I, the creator of the storyboard and collection was trying to say through it."

"I realized I wasn't sure how exactly the connections should work. Was I to connect the authors to their original texts first and then connect the author to another author? Was it better to connect one text to another and then connect the authors together through their texts? I realized halfway through that the different platforms I first found important to link the texts became less and less important but I didn't feel that I could take them out completely. Perhaps I was missing the ability to map out these connections in a diagram. Was this something that you needed talent or training to do effectively? In the same way that there are some people that are better at describing things in details while others are better at summarizing? Was there a block in my ability to figure out how to express the relationships in a screenshot? The alternative was that I had chosen too complex of a relationship to depict. Whatever the case I felt that I was in need of more time spent just learning how to build these types of relationships in such a style to produce something that could be deemed 'useful' for research practices."

"I was also curious as to how one goes about expressing the relationships between authors and books without having a way to show how time can alter them. Take my storyboard for example. What happens later down the road where Nicholas Carr decides to write a book that refutes his past reservations about technology? Or perhaps Craig Mod once again changes his stance on using the iPad as a platform for books? My storyboard would become practically useless. I thought of the enormous possibilities of RoSE if it was able to express the shifts of relationships between authors, texts and ideas. It seemed much more of a useful tool for a researcher who wanted to see how the relationship had changed in how these authors viewed distraction/hyper-attention as we progressed into the next 20 years. Without such a function in which one might "play" the change of relationships over time it feels that the project has a very static view of how the texts, authors and relationships work. This unfortunately goes against how social networks truly interact. Relationships are built, crumble, rebuilt or become stronger or weaker over the course of time. I think the RoSE project could be much more qualified as a "research tool" with the adding of such a function."

"My largest frustration came from the flaws in the program itself. Not being able to save my storyboard as I progressed; not being able to add things to my collection while working in storyboard mode; having to start over if I accidentally deleted a node that I needed, which funny enough happened in my final attempt to place this project together at which time I decided to draw in Cathy Davidson myself."

"I also found that perhaps the ability to expand the size of your storyboard would also be more helpful for times when those dealing with a very complex grouping of relationships need it."

█████████:

"Although I'm fairly certain this will change as the site develops more, I was expecting more social networking functions for this kind of project. Being able to add another member to a collection is a good start in that direction, but it's not quite as straightforward as a friends list or directory of contacts."

"I had fun tagging documents and people with keywords, but without a bank of already-used terms to pull from, I had trouble deciding among descriptors."

"One issue I believe we addressed in class is that the workspace is actually a little small, which is a minor hindrance on the storyboard."

"With the sheer amount of data that has the potential to be entered into the system, there is an equally large potential for the entry of incorrect or outdated information…. The alternative would be that RoSE users simply confirm this data by their own methods, but that feels like an essential element of research is being omitted from what should be a complete experience from RoSE (by which I mean that RoSE should ultimately become a website that you go to at the start of your research, and use all the way through to completion, having to use as little outside resources as possible)."

█████████:

1. when adding items, no matter what type of document you choose the program defaults to 'book.' In order to fix this, you must then navigate to the document itself through the 'documents' page and edit from a drop down menu.

2. the text prompting device in 'keywords' only considers the first 7 letters or so. Typing 'American' produces a drop down menu of the first 30 keywords but when you press the 'next' button for the next 30 it goes to page 1 of all keywords. Typing anything after 'American', 'American Poet', for example, produces no extant options.

3. the 'link to document' hyperlink field on the 'documents' fields does not save/work

4. too much redundancy. pick a place to start, say, person. After filling out name and any other pertinent, available information, have a link or dialog box at the bottom asking, "Would you like to add a document/keyword/photo for so-and-so? That link would then send you to the document page pre-filled with pertinent information previously entered (for the author/creator).

5. not sure about mult. authors per document. second and further authors of the same document must be added by way of 'document-to-person' relationship. the closest option is 'author of'

6. more highly organized hierarchy of keywords

7. document to document relationships: when/after I create a document, I can't relate it to another. However, I can relate document-to-document from an existing document to my new one (albeit in backwards order in the case of the article "electronic reading devices…" printed in the LA Times, is reversed)

8. can't edit a document's publish date from documents page, it must be inputted from the 'ad items' page

9. 'go to collection' button works but clicking 'storyboard for …' from collections page opens an empty collection box. instead, you have to go to your home page, then click storyboard for [name of your home page]; collections will then be displayed on the left

10. Icon size/style needs more options (in storyboard)

11. seems like something that could be adapted to a 3d rendering environment where a sphere can be rotated, then augmented

12. I have delete permissions on 'keywords' and it indicates that I made keywords that already existed (but I didn't, and had used them during a previous session). On the keywords themselves, there are no documents attached but the individual document pages will list the same keywords.

13. more on keywords: go to person page. type 'media' in keyword field. first name is Steve Agee. his keywords are: actor, American television actors, celebrity, coMEDIAn, writer. Same goes for anyone who is a MEDIAtor. This is not necessarily bad, but this difference should be explicated for each entry match. In other words, each keyword matched document or person should display exactly how that match was made, otherwise you have to go one by one.

14. anyone who is using the network to compile a list of references is going to want to be able to readily see what they've compiled so far, rather than continually visit the collection page. The finished interface should have a small vertically scrolling window (say, top right) that instantly updates and displays the documents, people, keywords, etc. selected for that particular collection.

15. when a user has more than 1 collection, each new page navigated to default-resets to the first collection. So, if I'm compiling my 3rd collection, I have to select that 3rd collection every time I want to add an item to it, even if it has been the only collection I've been working on that session.

16. the back button function could be more intuitive; something like Ebay's 'Go back to search results' button would be helpful.

██████████:

"While a valuable visualization, I had to spend more time fiddling with the logistics of the feature than on building actual content. I think the biggest flaw in the feature currently is that, since it cannot be saved, you cannot add addition nodes once you have begun building a storyboard. This proved troublesome for me, and I had to redo the project several times when I realized that I need to add an additional keyword. The easiest method I found to circumvent this problem was Paint. After saving my screen-shot in MS Paint, I could copy the color and shape of the node I wanted and paste it on the workspace to create additional nodes that RoSE wouldn't let me add."

"Another aspect of the storyboard that I struggled with was the size of the workspace itself. I think it would be beneficial to have the ability to "zoom in" or "zoom out," allowing for the distance in between the nodes to have significance."

"Additionally, I wished I could alter the size of the nodes— for instance, making the strong connections larger and the minor connections physically smaller."

▮▮▮▮▮▮▮▮:

"Also, making the entries themselves more substantial would be an improvement. Maybe making it mandatory to provide a link to articles to prove that it actually exists and is a reputable document."

▮▮▮▮▮▮▮▮:

"I was rather disheartened by the fact that one cannot turn the arrows into connectors. The solution would be that anything the creator can make with the 'draw' function should also carry with it the option to attach nodes to that line, or rectangle, or arrow, or text box even."

"I also found aspects of navigation within the site somewhat frustrating. It is possible that I was just moving around inefficiently, but sometimes switching between collections and searching for people and documents felt cumbersome. I felt like it would have been easier to employ a drag-and-drop function between people and collections. I just kept getting so frustrated after creating a person or document and then not being able to add it with ease to my collection. Sometimes, although only once really, did I begin to type in a pre-existing person or document to be added to my collection and the database found it within its archives. That inconsistency in the search function should be branched out so that searching more accurately examines the entire site for possible entries rather than for exactly matching words."

"One last thing though, I had problems adjusting which colors I wanted to use in my storyboard. Maybe more color choices while selecting?"

▮▮▮▮▮▮▮▮:

"I think the biggest limitation associated with RoSE will be whether or not there is someone overseeing the validity of the information that is being added to the databases. It would be easy for someone to add false information to RoSE without some kind of moderation on the users actions."

"While it was useful for my own personal understanding, I feel that it [the story board feature] would be difficult for any other viewer of the board to understand the information it is presenting without a detailed analysis to accompany it. One of the most challenging aspects was trying to fit the sheer volume of information I wanted to portray onto such a small space. The size of the nodes cannot be changed, and the text runs over itself, making it hard to read at times, and requiring very specific placement, which may not be consistent with how I would like the visualization to look."

▮▮▮▮▮▮▮▮:

"While, I really enjoyed the process and opportunity to visualize my works with the storyboard, I would say that the current storyboard application is somewhat limiting in its abilities, particularly considering spatial limitations. Someone in class suggested a zoom option, which would allow the user to get closer for smaller visualizations or zoom out for larger diagrams, and I think that would be extremely beneficial. While creating the storyboard, I found myself struggling to construct ideas and compose shapes in a manner that would both fit within the space provided and assist, rather than detract from, a new way of learning."